THE UNIVERSITY *of* EDINBURGH

Integrated structural biology approaches for the study of nucleic acid binding proteins

Silvia Visentin

Doctor of Philosophy

The University of Edinburgh

2020

Contents

Acknowledgements

I want to express my deep gratitude to my supervisor Dr Laura Spagnolo for giving me this opportunity and providing continuous support. Laura always believed in me and she gave me all the freedom every young scientist would like to have. Grazie Laura per tutte le chiacchierate e le risate che ci siamo fatte guardando le immagini di CSM e CMR (e pensando a cosa potessero assomigliare...); grazie per avermi dimostrato che è possibile trovare un'amica in un supervisor.

I would also like to thank Dr Sveta Makovets, Dr Luke Clifton, Professor David Bhella for their support and precious advice, not only during my assessments but also throughout my entire PhD. Thanks to Dr James Doutch and Dr Brian Smith for their patience and their continuous help, advice and constructive feedback.

A special thanks goes to Dr Giuseppe Cannone, not only for his never-ending help but also for being one of the best people I have ever met. Anche se abbiamo passato pochissimo tempo assieme nel lab, ricorderò per sempre tutti i preziosi consigli che mi hai dato. Grazie Beppino per la tua amicizia, per le lunghe telefonate e per aver sfidato il traffico per un veloce abbraccio.

I want to thank all the fantastic people I met in Scotland and who will always be in my heart. A special thanks goes to Gabbro, Sheila & Francesco, my Italian family abroad. I also want to thank all the pure brilliant people from the lab (Chantal, Caroline, Abbie, Éilís, Will, Connor, Martin, Dima and Louie) for the coffee breaks and the Fridays at the pub.

Grazie alla mia famiglia per il loro amore e continuo supporto. Grazie papà e mamma per avermi dato non solo la vita, ma anche la forza e il coraggio di viverla a pieno.

I am grateful to the School of Biological Sciences and ISIS Neutron and Muon Source for funding my PhD and without whom this thesis would not have been possible.

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Silvia Visentin

Lay Summary

Proteins are large size molecules (macromolecules) that are involved in every aspect of life. Every protein has a specific job to accomplish in the cell. The shape (structure) that proteins adopt depends on the tasks they have to perform. Most proteins can bind many different molecules and for this reason it cannot have a rigid structure but it has to be able to shift from a configuration to another. Therefore, proteins are dynamic macromolecules and studying the structural basis of these movements is important to understand how they work.

Until recently, two main techniques, called X-ray crystallography and nuclear magnetic resonance (NMR), were used to determine the structure of proteins. However, the use of these “traditional” methods in isolation has some limitations. For example, X-ray crystallography can be used to resolve the structure of compact proteins and the three-dimensional reconstruction obtained only represent a snapshot of the protein in a particular state. On the other hand, nuclear magnetic resonance allows to obtain a series of structures, each of them representing an individual conformation adopted by the protein, therefore allowing the analysis of biologically relevant protein flexibility. However, NMR has the disadvantage that it can be used only for relatively small proteins.

Nowadays, the efforts of the scientific community has led to the development of new techniques that can be combined in order to allow the structural characterization of “difficult” proteins, meaning those that cannot be characterized with traditional methods. Gathering structural information from different methods helps to obtain a more complete representation of the conformations adopted by proteins, not only in isolation but also as part of a biological network. In this study, the combination of several techniques was used to study the shape of Staufen: a small, flexible protein, which is difficult to study using only conventional methods. Additionally, this strategy is also beneficial to understand the dynamic behavior of large macromolecular complexes, such as the one formed by the protein MCM. Understanding the structure of proteins allows to uncover the molecular basis of their functions, not only in healthy cells but also in abnormal ones. Consequently, this knowledge is necessary to design “smart” drugs that can be used in the treatment of diseases, such as cancer.

Abstract

Proteins are the ultimate effectors of biological mechanisms and are involved in every aspect of cellular life. The functional properties of proteins depend on their three-dimensional structure. Indeed, proteins are not rigid entities and internal motions, on a wide range of timescales and distances, are necessary to accomplish a specific function. While certain proteins adopt a compact conformation and undergo small-scale rearrangements, others are more flexible and withstand more dramatic movements. Proteins that exhibit a regulatory role function by binding multiple partners, such as small molecules, DNA, RNA, other proteins. Increased levels of flexibility favour this binding promiscuity.

A long-standing goal in molecular biology has been the development of new methods to enable the determination of three-dimensional structures of proteins that exhibit a high level of dynamic complexity. In fact, it is becoming clearer every day that individual structural techniques have several limitations. An integrative structural biology approach provides the tools to decipher the dynamic configuration of proteins by combining information from multiple sources, including biochemical, bioinformatics and biophysical methods.

In this thesis, an integrated approach is used to structurally characterize two extremely different nucleic acids binding proteins, the human Staufen1 protein and the archaeal *Pyrococcus abyssi* MCM complex. In this study, I show that an integrated structural biology approach is not only essential to determine the architecture of small, flexible proteins, which are traditionally difficult targets for conventional approaches, but it is also beneficial to understand the dynamic behavior of large, more compact macromolecular complexes.

List of Figures

List of Tables

Abbreviations

AAA+	ATPases associated with diverse cellular activities superfamily
ADP	Adenosine diphosphate
Arf1	ADP-ribosylation factor 1
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
BN-PAGE	Blue native-polyacrylamide gel electrophoresis
BRET	Bioluminescence resonance energy transfer
Cdc45	Cell division cycle 45 protein
Cdc6	Cell division cycle 6 protein
CDK	Cyclin-dependent kinase protein
Cdt1	Chromatin licensing and DNA replication factor 1 protein
CMG	Cdc45-MCM-GINS complex
CTF	Contrast transfer function
DBD	DNA-binding domain
DKK	Dickkopf protein
DNA	Deoxyribonucleic acid
dsDNA	Double stranded deoxyribonucleic acid
dsRNA	Double stranded ribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
eIF2	Eukaryotic initiation factor 2 protein
EM	Electron microscopy
EOM	Ensemble optimization method
FID	Free induction decay
FSC	Fourier shell correlation
GINS	*go-ichi-ni-san* complex (Sld5-Psf1-Psf2-Psf3 complex)
HD	Homeodomain
hiCLIP	RNA hybrid and individual-nucleotide resolution ultraviolet crosslinking and immunoprecipitation
HMG	High-mobility group box proteins
HPLC	High performance liquid chromatography

HSQC	Heteronuclear single quantum correlation
HTH	Helix-turn-helix
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IR	Infrared
ITC	Isothermal titration calorimetry
KH	K homology
LB	Lysogeny Broth (Luria-Bertani Broth)
MALLS	Multi-angle laser light scattering
MCM	Mini-chromosome maintenance protein
mRNA	Messenger ribonucleic acid
MST	Microscale thermophoresis
MWCO	Molecular weight cut-off
NMR	Nuclear magnetic resonance
ORC	Origin recognition complex
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PKR	Protein kinase R
pre-IC	Pre-initiation complex
pre-RC	Pre-replication complex
PTM	Post-translational modifications
RBD	RNA-binding domain
RBP	RNA-binding protein
RNA	Ribonucleic acid
RNAPII	RNA polymerase II protein
RRM	RNA recognition motif
RT	Room temperature
SAS	Small angle scattering
SANS	Small angle neutron scattering
SAXS	Small angle X-ray scattering
SBS	Staufen binding site
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography

SMD	Staufen-mediated decay
SPA	Single-particle analysis
SPR	Surface plasmon resonance
ssDNA	Single stranded deoxyribonucleic acid
SSM	Staufen swapping motif
Stau	Staufen protein
TB	Terrific Broth
TBD	Tubulin binding domain
TBE	Tris/Borate/EDTA
TEM	Transmission electron microscopy
TMS	Tetramethylsilane
TROSY	Transverse relaxation optimized spectroscopy
Upf1	Up-frameshift suppressor 1 (Regulator of nonsense transcripts 1) protein
UTR	Untranslated region
UV	Ultraviolet
UV-Vis	Ultraviolet-Visible
ZF	Zinc finger

Chapter 1

Introduction

1.1 The evolutionary advantage of multidomain proteins

Proteins mediate every aspect in life. They are made of domains, which represent their fundamental building blocks. A domain can be defined as an independent evolutionary unit (single-domain protein) or be a module of a larger construct (multidomain proteins). Domains are not only the structural elements of proteins, but also their functional components as they often retain an independent function or act in cooperation with other domains to contribute to the function of a multidomain protein (Vogel et al., 2004). The majority of proteins are formed of at least two domains, whose nature and interactions govern the function of the protein. In multidomain proteins, the connecting linkers play an essential role in maintaining cooperative inter-domain interactions (Gokhale and Khosla, 2000) and display different degrees of flexibility to perform their particular biological purpose. However, loops and linkers do not act only as flexible connectors within and between distinct domains of a protein, but they have an important role in protein function and allostery (Papaleo et al., 2016). The formation of new domain combinations, in a well-defined architecture governed by linkers, is an important mechanism in protein evolution (Vogel et al., 2004) and it allows to evolve new functions.

1.2 Flexibility as a functional “tool”

Protein flexibility is essential for biochemical function and it allows proteins to adapt, and subsequently bind, to their target ligands. However, not all proteins are equally flexible and the degree of flexibility varies between proteins. For example,

some proteins behave as they were relatively rigid or compact. Conversely, many proteins undergo very large conformational changes, for example upon binding to a substrate or a ligand. In particular, the higher degree of flexibility of regulatory proteins, when compared to other functional categories, has been shown to relate to increased binding promiscuity allowing specificity towards multiple binding partners (Liu et al., 2009). Structural and proteomic research over the past decade has expanded the traditional structure-function paradigm by establishing the functional relevance of protein dynamics (Pabon and Camacho, 2017). In this perspective, particular efforts have been put towards the study of intrinsically disordered proteins (IDPs) or regions (IDRs), which exhibit large structural rearrangements, including the formation or loss of secondary structure elements, triggered by environmental changes or interactions with partners (Craveur et al., 2015). Functional and evolutionary flexibility in protein-protein interaction networks are considerably enhanced by IDPs and IDRs, as their interaction profile can be fine-tuned with alternative splicing and post-translational modifications (PTMs), which may result in allostery and promiscuity (Tompa et al., 2015). The continuously increasing amount of research in this field is motivated by the involvement of IDPs and IDRs in several crucial biological functions (Dunker et al., 2000, Dunker and Obradovic, 2001). In fact, intrinsically disordered proteins or regions of proteins play key roles in regulatory and signalling processes of the cell (Wright and Dyson, 2015, Tompa, 2011).

1.3 Protein complexes

Every biological process requires the assembly of individual proteins into functional complexes. In particular, molecular machines are complexes that perform the major biological processes in the cell and consist of multiple components, whose activities are largely sequential and independent. These units retain their individual functions, but act in a more coordinate way because they are part of the same complex. Therefore, the formation of protein complexes is evolutionary favourable as it allows to bring together diverse functions by reducing the additional regulatory potential that would be required if the different polypeptide chains were not part of a complex (Marsh and Teichmann, 2015).

Additionally, proteins can self-assemble into homomeric complexes. In this case, the evolutionary benefit is obviously not an increase in the number of concerted different activities and several functional advantages linked with homo-multimerization have been suggested. Self-assembly is directly related to protein function because the interface between two monomers could generate an active site or a ligand-binding site, it is a simple mechanism to achieve cooperative function and increased protein stability and it is "economically" favourable for the cell because it allows to form large assemblies using less genetic material (Perica et al., 2012, Goodsell and Olson, 2000). For what concerns the dynamics of protein complexes, the conformational changes observed upon protein-protein interactions largely depend on the nature of individual components. In general, if a protein exhibits a high degree of flexibility as a monomer, it will undergo a large conformational change upon binding. An extreme example of this phenomenon is represented by intrinsically disordered proteins, which have been shown to be able to form highly ordered structures when part of a complex (Hegyi et al., 2007, Dyson and Wright, 2002). As previously discussed for monomeric proteins, also protein complexes present movements on a wide range of time-scales and locations. In several energy-dependent molecular machines, large coordinated motions of different subunits are necessary for carrying out a specific function. For example, the rotary movement of the stalk of ATP synthase, caused by the movement of protons through the membrane-bound region, induces a conformational change of the β-subunits, which triggers ATP synthesis from ADP and phosphate (Stock et al., 1999). A different category of dynamic complexes is represented by "fuzzy complexes", which are formed by intrinsically disordered proteins that remain partially unfolded (Tompa and Fuxreiter, 2008).

1.4 Nucleic acid-binding proteins

Nucleic acid-binding proteins are fundamental for the life of every organism. DNA-binding proteins (DBPs) are key effectors of all aspects of genetic activity in the cell, such as transcription, packaging, rearrangement, replication and repair. RNA-binding proteins (RBPs) are involved in splicing, polyadenylation, RNA stability, RNA localization and protein translation. The physical interactions

between proteins and nucleic acids rely on hydrogen bonds, hydrophobic interactions and water-mediated contacts. Many DBPs and RBPs contain multiple nucleic acid binding domains, which can act together to recognize different regions of their targets, leading to an increase in affinity and specificity.

1.4.1 DNA-binding proteins

DNA-binding proteins (DBPs) bind DNA using a wide range of DNA binding domains (DBDs), such as homeodomain (HD), helix-turn-helix (HTH), zinc finger (ZF) and high-mobility group box (HMG). The mechanisms by which DBPs bind their targets are historically known as “direct readout” and “indirect readout” (Otwinowski et al., 1988). “Direct readout”, also known as “base readout”, occurs when the side chains of the amino acids of a protein directly interact with specific bases of the DNA. On the other hand, “indirect readout” is triggered by specific conformation of the DNA, such as DNA bending and kinks, and it is therefore referred as “shape readout”. Even though direct recognition, through a conserved DBD, is the common end-point in a protein-DNA complex, their interaction can depend on other factors. For example, DNA binding can be mediated by protein disordered regions. This mechanism is adopted by many transcription factors in order to facilitate protein diffusion along the DNA and regulate the transition from non-specific to specific binding (Doucleff and Clore, 2008). Moreover, DNA recognition can be mediated by cooperative binding when protein-protein interactions between multiple polypeptide chains are required to alter or stabilize protein-DNA contacts (Slattery et al., 2011).

1.4.2 RNA-binding proteins

RNA-binding proteins (RBPs) are generally considered as proteins that bind RNA with high affinity and specificity, through one or more globular RNA-binding domains (RBDs) (Lunde et al., 2007). Moreover, it is a common idea that RBPs interact with RNA targets in order to modify their fate or function in the cell (Gebauer et al., 2012). Even though these statements are true for many RBPs,

recent proteomic analysis have uncovered a variety of proteins that lack conventional RBDs (Tsvetanova et al., 2010, Castello et al., 2012, Baltz et al., 2012), such as RNA recognition motif (RRM), hnRNP K homology domain (KH) and DEAD box helicase domain, but are able to bind RNA using different strategies. RNA recognition can be mediated by shape complementarity. An example of this binding mechanism is represented by the internal ribosome entry site of hepatitis C virus, which adopts a complex “key-lock” interaction mode to interact with the 40S ribosome (Quade et al., 2015). RNA binding can also be mediated by protein disordered regions, which are highly enriched in RBPs and conserved from yeast to humans (Beckmann et al., 2015). Consequently, IDRs could represent plastic, multifunctional RNA-binding motifs, which are involved in both non-selective and highly specific protein-RNA interactions, often by triggering co-folding (Varadi et al., 2015). An example of this binding mechanism is represented by the intrinsically disordered Arg-Gly-Gly repeat motif of fragile X mental retardation protein, which co-folds with its RNA target resulting in a β-turn that recognizes a G-quadruplex RNA (Vasilyev et al., 2015, Phan et al., 2011). The variety of protein-RNA interactions reported suggests that RNAs may also regulate the function of proteins. It follows that there must be a functional cross-talk between proteins and RNAs, which allows both macromolecules to regulate each other’s fate (Hentze et al., 2018, Beckmann et al., 2016). For example, the antiviral protein PKR is activated upon binding to double stranded RNAs (dsRNAs) produced during viral replication, and subsequently phosphorylates the subunit α of the eukaryotic translation initiation factor eIF2, severely inhibiting cellular and viral protein synthesis (Balachandran and Barber, 2007).

1.5 The importance of structural biology

A fundamental axiom of biology is that the three-dimensional structure of a protein determines its function. Understanding function through structure is a primary goal of structural biology. In particular, the structural study of how proteins interact with DNA, RNA, small molecules and other proteins is necessary to gain a deeper knowledge of all biological processes. Many factors can impact the correct

performance of molecular machines in the cell and lead to the development of diseases. Deciphering the results of these changes on the physiological structure of proteins and their complexes helps to reveal the role they play in diseases. Consequently, structural biology is of extreme importance for other fields, such as biotechnology, drug discovery and the design of “intelligent” therapeutics.

1.6 Limitations of conventional structural biology techniques

The vast majority of the structures deposited in the Protein Data Bank (PDB) have been solved by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Therefore, we can refer to them as “conventional” structural biology techniques. Electron microscopy (EM) is the third major method represented in the PDB, followed by hybrid approaches. The continuous increase in the number of deposited structures solved by EM studies and hybrid methods is a clear indication of how the field of structural biology is evolving. In fact, even though X-ray crystallography and NMR are extremely powerful techniques that allow obtaining atomic resolution structures and are still widely used, they present several limitations that are being overcome by embracing other methods. The main disadvantages of X-ray crystallography consist in the difficulty of studying flexible proteins and in the impossibility of looking at dynamics. In fact, increasing flexibility generally corresponds to more difficult crystallization and, when this is achieved, to lower resolution crystal structures or “invisible” linkers or missing parts of the proteins (Yu et al., 2000). The strategies that are usually adopted to overcome these issues consist in producing polypeptide constructs depleted of extremely flexible portions and in mutating particular residues in order to enhance the rigidity of a protein or “lock” it in a specific conformation. Moreover, a crystal structure represents a “snapshot” of the protein in the particular configuration that is adopted in the crystal lattice, therefore in a non-physiological environment. On the other hand, in NMR experiments the protein is in solution and this allows studying its structure in more physiological conditions. Moreover, using this technique, it is possible to obtain an ensemble of the conformations that the protein adopts, allowing the characterization of flexible and dynamic systems. However, the limiting factor of NMR studies is the size of the protein or complex

of interest. In fact, the upper size limit corresponds to ~40 kDa and this comes from the fact that the width of resonances increases with molecular weight, causing assignment to become impossible because of spectral overlaps (Campbell, 2002).

1.7 The power of integrative structural biology

A deeper understanding of the dynamic nature of macromolecular machines and their role in biological processes can be obtained through an integrative structural biology approach. As discussed in the previous section, the number of structures deposited in the PBD, which have been solved by EM studies and hybrid methods, has been continuously increasing in the last few years. The characterization of more complicated systems, which cannot be studied solely by conventional techniques, requires a multi-angle approach that provides information at different levels of detail. Employing several biochemical, bioinformatical and biophysical tools allows to obtain an all-round model of the system of interest. Over the last few years, integrative structural biology has been able to model the architecture of extremely important complexes, whose structural determination seemed impossible by adopting individual techniques. A striking example of the success of this approach is the RNA polymerase II (RNAPII) transcription initiation machinery, whose architecture has been resolved by using small angle X-ray scattering, cryo-EM, cross-linking mass spectrometry, affinity pull-downs, labelling and co-localization experiments, combined with atomistic structural information in the form of homology models or crystal structures available for some of the domains (Trnka et al., 2019). Another example is represented by the aerolysin pore formation complex, where a hybrid method combining cryo-EM, X-ray crystallography and molecular dynamics analysis showed the transition from monomer to a heptameric pre-pore state and finally to a functional toxin pore inserted in the target membrane (Degiacomi et al., 2013). These two examples represent a clear indication that the structure of macromolecular machines and their dynamic behaviour in biological processes can be studied using integrative structural biology.

In this thesis, an integrated approach is used to structurally characterize two extremely different nucleic acids binding proteins, the human Staufen1 protein and the archaeal *Pyrococcus abyssi* MCM complex.

Staufen1 is a dsRNA-binding protein involved in many aspects of RNA regulation, such as mRNA transport along the microtubules (Vessey et al., 2008, Martel et al., 2010, Brendza et al., 2000, Ramasamy et al., 2006, Martel et al., 2006), mRNA decay (Park, 2013, Park and Maquat, 2013, Kim et al., 2005, Kim et al., 2014, Gong et al., 2013, Gong and Maquat, 2011, Gong et al., 2009, Cho et al., 2012, Cho et al., 2013) and regulation of protein translation (Ravel-Chapuis et al., 2012, Elbarbary et al., 2013, Bonnet-Magnaval et al., 2016).

MCM is a highly conserved DNA helicase complex involved in replication (Tye, 1999, Costa et al., 2008, Costa and Onesti, 2008), genome repair (Bailis and Forsburg, 2004) and transcriptional regulation (Hubbi et al., 2011). Archaeal MCMs have proven to be powerful tool to elucidate essential features of the functions of this protein in Eukaryotes, thanks to their simplified structural organization (Miller and Enemark, 2015).

A more detailed introduction to the role of Staufen1 and MCM proteins and on their structural characterization can be found in Chapter 3 and Chapter 4, respectively.

These two proteins are used in this study to show how an integrated structural biology approach is not only essential to determine the architecture of small, flexible proteins, which are traditionally difficult targets for conventional techniques, but it is also beneficial to understand the dynamic behaviour of large, more compact macromolecular complexes.

Chapter 2

Biophysical methods

2.1 Size exclusion chromatography coupled with multi-angle laser light scattering

Size exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALLS) is a technique widely used for protein characterization. SEC-MALLS allows to evaluate sample purity and stability, as well as to determine proteins oligomeric state during purification (monomers, dimers, trimers etc.) and their size.

The physical phenomena at the basis of SEC-MALLS is that small species do not present variation in their scattering intensity and are therefore referred to as “isotropic scatterers”. On the other hand, the light scattered by big biomolecules is different in every direction and its intensity depends on the scattering angle. It follows that detectors located at different angles perceive different amount of light scattering and this can be used to calculate the molar mass of the species.

In order to perform a SEC-MALLS experiment, an HPLC system equipped with a size exclusion chromatography (SEC) column is connected to a multi-angle light scattering detector and to a differential refractive index detector (Figure 2.1). The sample is loaded onto the SEC column and subsequently passes through the MALLS detector. The resulting chromatogram is complemented with the molar mass distribution of the sample, allowing to determine the exact molecular weight of the species in a particular peak. Consequently, SEC-MALLS allows to overcome conflicting results from size exclusion chromatography alone. Indeed, in a SEC experiment, the separation is based on the hydrodynamic size of the biomolecules and this could resolve in an erroneous interpretation of the oligomeric state in solution.

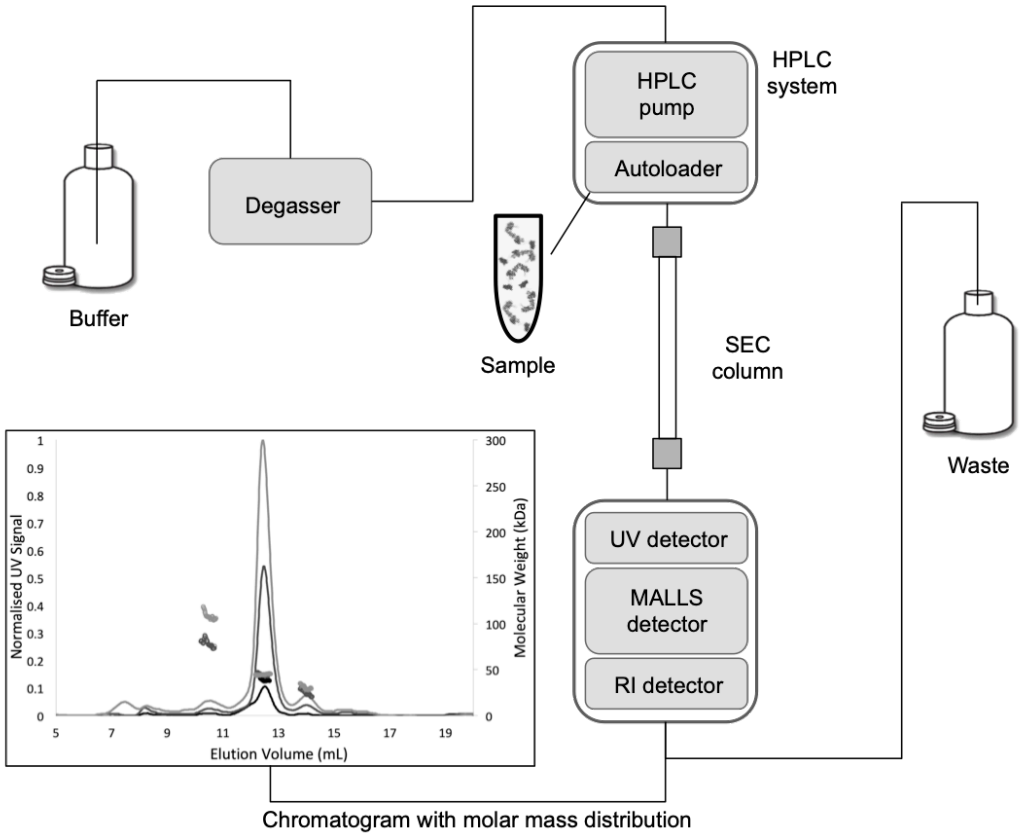


Figure 2.1: Schematic representation of SEC-MALLS experiment. A HPLC system equipped with a size exclusion chromatography (SEC) column is connected to a UV detector, a multi-angle light scattering detector and a differential refractive index detector. The sample is loaded onto the SEC column and subsequently passes through the detectors. The resulting chromatogram is complemented with the molar mass distribution of the sample, which allows to determine the exact molecular weight of the species in a particular peak.

2.2 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a powerful technique that allows to determine the molecular weight, hydrodynamic and thermodynamic properties of a protein or other macromolecules in solution. Moreover, it allows to study protein interactions, such as oligomerization states and complexes formation, their stoichiometry and association constants.

An analytical ultracentrifuge is a high-speed centrifuge equipped with optical systems (UV-Vis absorbance and Rayleigh interference) that enable to record the behaviour of molecules in solution as they undergo sedimentation (Figure 2.2). As the rotor spins, scans are generated when the cell, containing the sample, passes through an optical detector's path. Sequential scans allow to monitor the change in local concentration of a protein in solution and therefore its "radial position".

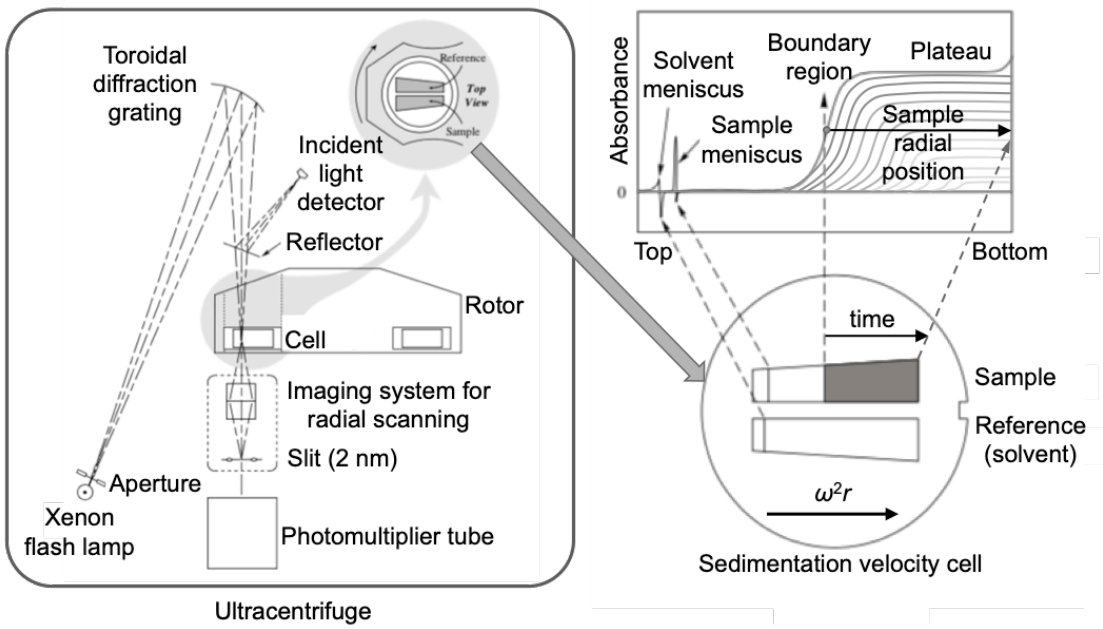


Figure 2.2: Schematic representation of AUC system and sedimentation velocity experiment. Left: High-speed centrifuge equipped with optical systems (a Xenon flash lamp generates a high-intensity light that is directed to the cell by a toroidal diffraction grating; light is diffracted by inhomogeneity in the cell and recorded by an imaging system connected to a photomultiplier tube). Right: Sedimentation velocity experiment showing the movement of macromolecules upon sedimentation, represented by the shift of boundaries, from the top to the bottom of the cell. (Adapted from [(Ralston, 2004)])

There are two approaches that can be used in analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. Sedimentation velocity is technique in which the macromolecules of interest sediment during the duration of the experiment, ultimately pelleting at the bottom of the cell. On the other hand, in a sedimentation equilibrium experiment the particles reach a perfect equilibrium between sedimentation and diffusion and do not sediment.

Thanks to the shorter duration of the experiment and to the development of softwares that allow to model complex systems, sedimentation velocity AUC is generally considered the primary method of choice for most current applications of analytical ultracentrifugation, including the analysis of protein oligomerization state and hetero-complexes formation, conventional targets of sedimentation equilibrium AUC experiments.

In a sedimentation velocity experiment, a series of sequential scans is used to monitor in real-time how the macromolecules move through the solution, thanks to snapshots that describe their position in the cell in a specific moment in time.
The rate of sedimentation of the particles provides information about their sedimentation coefficient *s*, which is described by the Svedberg equation as:

$$s = \frac{M(1 - \bar{\nu}\rho)}{f}$$

where *M* is the molar mass of the macromolecule, $\bar{\nu}$ its partial specific volume, ρ the density of the solvent and f the frictional coefficient.
It follows that the sedimentation coefficient is proportional to the size of the macromolecule and inversely proportional to the frictional coefficient, which is related to the overall dimensions of the particle. In other words, if two proteins have the same molar mass but different shapes, *i.e.* different frictional ratios, they will sediment at different Svedberg (S) units (10^{-13} seconds).

In an ideal mono-disperse system, the rate of sedimentation can be calculated by plotting the midpoint of each curve generated by individual scans, therefore monitoring the progression of the “boundary”. However, in real experimental condition, this simple approach is often not applicable because of the presence of extremely small molecules or of several species in solution, even when highly purified samples are employed. To overcome these limitations, several

computational approaches have been developed in order to analyse sedimentation data. For example, the programs generally used to analyse multi-species systems are UltraScan and Sedfit, which are based on the c(s) analysis and allow to obtain the sedimentation coefficient *s* (S). Moreover, these softwares allow to determine the steepness of the boundaries, which relates to the frictional ratio f/f_0 that represents the comparison between the frictional coefficient of the sedimenting particle (f) and the one of an ideal sphere of identic volume (f_0). It follows that it is possible to extrapolate the molecular weight of the species in solution by calculating the sedimentation coefficient and the frictional ratio.

2.3 Microscale thermophoresis

Microscale thermophoresis (MST) is a biophysical technique that allows to determine the strength of inter-molecular interactions based on the dispersion of biomolecules upon heating, in solution.

Thermophoresis is the movement of the molecules across a temperature gradient and it is affected by their size, charge and hydration shell. Consequently, the change in any of these properties, upon binding to a partner, can be monitored using this technique and it allows to determine the dissociation constant (K_d) of the complex.

In a MST experiment, an infrared (IR) laser is used to heat a small portion of a capillary tube containing the sample, as shown in Figure 2.3. Upon laser stimulation, molecules can move away from the heat (“positive” thermophoresis) or towards it (“negative” thermophoresis). In both cases, once the laser is turned off, the molecules diffuse back to their original equilibrium state. The difference in the molecular diffusion of a complex and of an unbound protein is measured as a change in fluorescence. Indeed, one of the biomolecules (“target”) is labelled with a fluorophore, while the other molecule is unlabelled and mixed to the target in different concentrations, acting as a “ligand”. Upon heating, if the ligand is interacting with the target molecule, the complex will diffuse at a different rate relatively to the individual components, giving rise to a difference in fluorescence (Figure 2.3, bottom left). The change in fluorescence amongst the capillaries,

containing different concentrations of ligand, allows to directly determine the dissociation constant of the complex (Figure 2.3, bottom right).

MST presents several advantages compared to the conventional techniques traditionally used to study kinetics. In contrast with Isothermal Titration Calorimetry (ITC) and Surface Plasmon Resonance (SPR), the amount of sample required is limited to few microliters. In fact, each capillary contains only 4-5 μl of sample and a 16-capillaries titration curve is generally sufficient to observe the complex formation and obtain a precise measure of the K_d. Moreover, this method does not require immobilization of the target protein and it is performed in solution, therefore in native-like conditions. Using this technique, any size of biomolecules can be studied and there are virtually no buffer limitations due to absence of buffer interference.

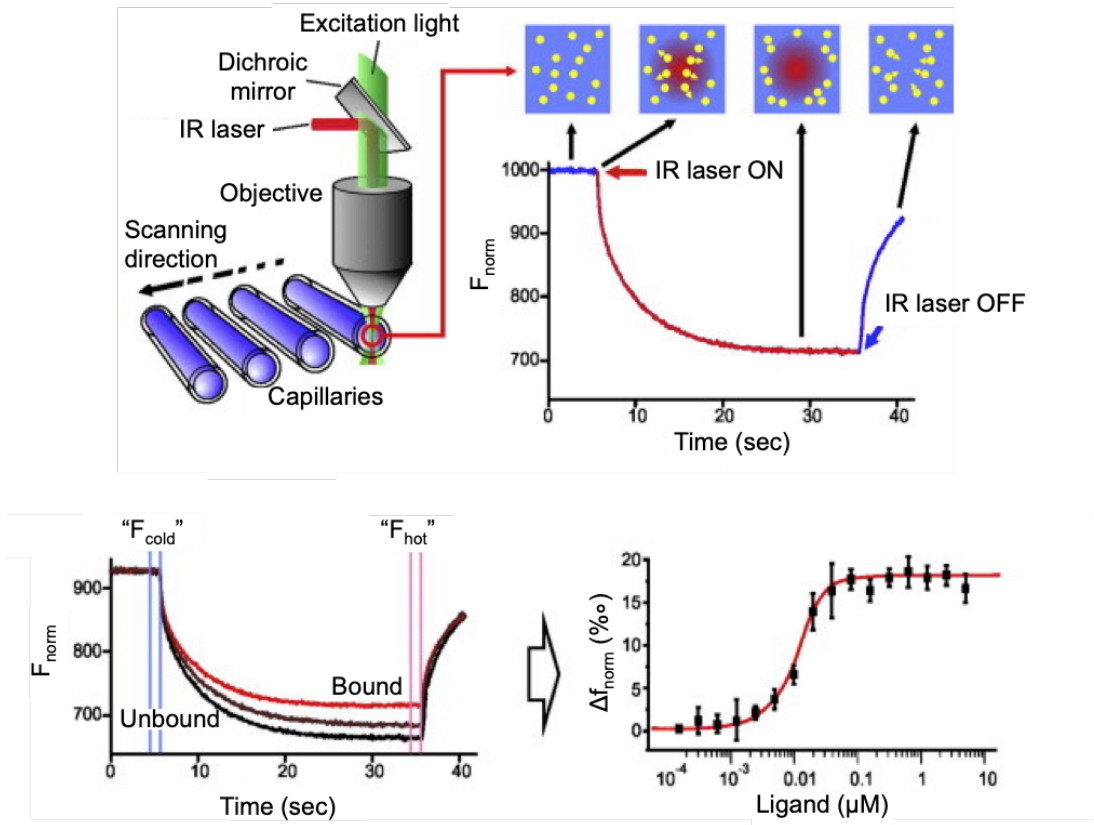


Figure 2.3: Schematic representation of Monolith Nanotemper system and principles of MST experiment. Top: IR laser is used to heat a small portion of a capillary tube containing the sample and the heat generated causes the molecules to move; once the laser is turned off, the molecules diffuse back to their original equilibrium state. Bottom: A complex will diffuse at a different rate when heated up relatively to the individual components, giving rise to a difference in fluorescence that allows to directly determine the dissociation constant of the complex. (Adapted from [(Jerabek-Willemsen et al., 2014)])

2.4 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance spectroscopy (NMR) is one of the techniques that has been traditionally used to study the three-dimensional structure of proteins at atomic resolution, together with x-ray crystallography. NMR allows to determine the solution structure of relatively small proteins or complexes, generally in the range of 5-40 kDa. However, this method can also be used to yield information about the chemical properties of macromolecules and their folding state, or to measure the kinetics of complex formation.

The NMR phenomenon is based on the fact that nuclei of some atoms (^{1}H, ^{13}C, ^{15}N, ...) have magnetic properties, which can be used to gain chemical information on the geometry of a molecule. Indeed, those atoms possess a “non-zero” overall spin because the number of the protons and neutrons in their nucleus is different.

When an external magnetic field B_0 is applied, the nuclear magnetic momentum of an NMR-active atom will align to it, either in an energetically favourable parallel orientation (with spin $m = +\frac{1}{2}$) or in an higher energy anti-parallel orientation (with spin $m = -\frac{1}{2}$) as shown in Figure 2.4 A.

However, the rotational axis of the spinning nucleus is not perfectly parallel (or anti-parallel) to the applied magnetic field B_0, but it rotates at an angle. The angular velocity, also called Larmor frequency, can be described as:

$$\omega_0 = \gamma B_0$$

where γ is the magnetogyric ratio and it is characteristic of each nucleus as it relates to its magnetic momentum μ and spin number I.

Conceptually, the NMR experiment consists in irradiating the sample with a pulse of electromagnetic radiation of frequency ω_0 (in the MHz range) and duration τ. In this way, a nucleus can be induced to transition from a low energy parallel orientation to a higher energy anti-parallel state and this absorption of energy is at the basis of the NMR method. It follows that, after the pulse, the sample nuclei magnetic moments are in a non-equilibrium condition, as they are not anymore aligned to the applied magnetic field B_0. Subsequently, the nuclei magnetic moments initiate a process called “relaxation”, in which they return to

equilibrium and therefore aligned to the external magnetic field. The time-dependent relaxation signal is recorded as a FID (free induction decay) spectrum and then Fourier transformed in order to obtain a frequency domain spectrum (Figure 2.4 B).

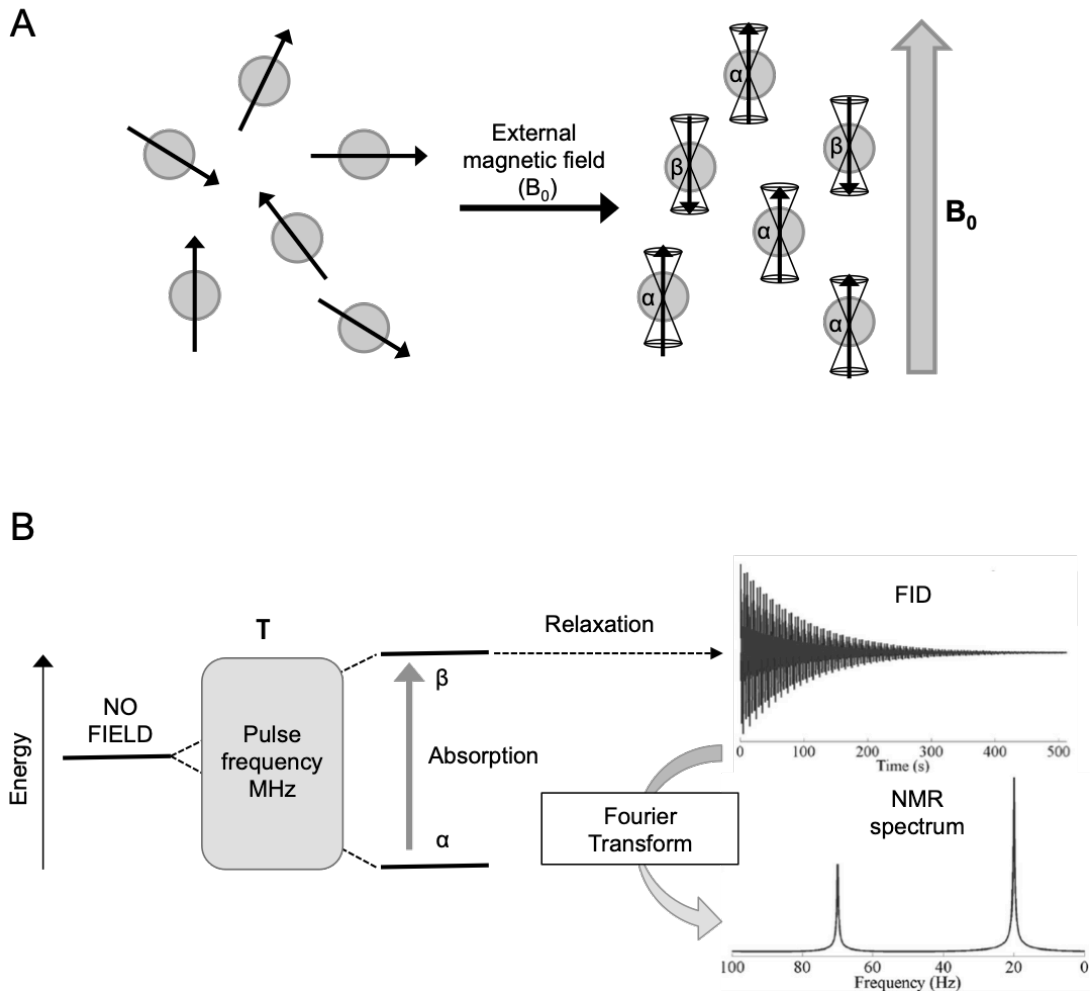


Figure 2.4: Schematic representation of NMR experiment. A: Schematic representation of the rearrangement of the nuclear magnetic momentum of NMR-active atoms upon application of magnetic field (spin m= ±1/2); the rotational axis of the spinning nucleus is not perfectly parallel (or anti-parallel) to the applied magnetic field B_0, but it rotates at an angle. B: The sample is irradiated with a pulse of electromagnetic radiation that induces the nuclei to transition from a low-energy to a high-energy state; this absorption of energy causes the nuclei magnetic moments to be in a non-equilibrium condition and subsequently to return to equilibrium *via* a process called "relaxation": the time-dependent relaxation signal is recorded as a FID (free induction decay) spectrum and then Fourier transformed in order to obtain a frequency domain spectrum.

The power of the NMR method is given by the fact that it allows to distinguish between chemically different protons, which experience a slightly different applied magnetic field depending on the electronic environment in which they are located. The electrons around the nuclei generate a magnetic field, which is anti-parallel to the external one and attenuates its effect on the atoms. It follows that protons are electronically shielded and the Larmor frequency can be re-written as:

$$\omega_0 = \gamma(B_0 - S)$$

where S represents the change in magnetic field caused by the electrons, and it depends on the chemical structure and geometry of the molecule.

The NMR spectra provides information on the "chemical shifts" of all the protons of the examined molecule, in comparison to the signal given by tetramethylsilane (TMS) or by 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), which are used as reference standards for NMR experiments of organic solvents or proteins, respectively.

When analysing proteins by NMR it is useful to introduce one or more NMR-active atoms, such as ^{15}N and ^{13}C. The labelled recombinant protein can be expressed in micro-organisms grown in minimal growth medium supplemented either with ^{15}N-labelled ammonium salts or with ^{13}C-labelled glucose. The advantage of using heteronuclear NMR analysis is that the nuclei spins are almost all connected by one-bond couplings. In particular, the ^{1}H-^{15}N coupling is the most important starting point for NMR analysis of proteins since this bond is present in every amino acid residue, except for N-terminus and proline residues.

The simplest method used to record this coupling is called ^{1}H-^{15}N HSQC (heteronuclear single quantum correlation). However, when analysing larger proteins or complexes, it is generally preferred to perform a TROSY experiment (transverse relaxation optimized spectroscopy) in order to obtain sharper peaks in the NMR spectra and higher resolution.

2.5 Small angle scattering

Small angle scattering (SAS) is a powerful technique that has been extensively used to structurally characterize biological systems over a broad range of macromolecular sizes, from small peptides to large macromolecular assemblies. This method can be used on its own to obtain low-resolution models (10-20 Å) of biological macromolecules in solution, however it is more often used in combination with other structural, hydrodynamic, computational and biochemical techniques.

A schematic representation of a scattering experiment is illustrated in Figure 2.5. It is conceptually very simple, with many instruments being essentially pinhole cameras, and requires only a source, a sample and a detector, combined with focussing and collimation devices, such as mirrors and slits. Scattering is a physical process where a beam of radiation, in this case either X-rays or neutrons, is deviated from its initial incident trajectory by inhomogeneities in the medium through which it passes.

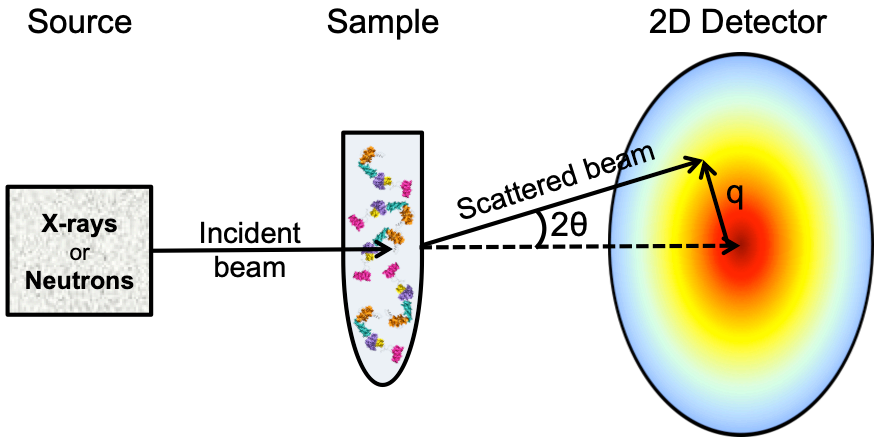


Figure 2.5: Schematic representation of SAS experiment. A beam of radiation is produced by a source and deviated by its incident trajectory by inhomogeneities of the sample; the scattered beam hits the detector and the scattering only depends on the momentum transfer *q* (scattering vector).

X-rays and neutrons are both used as probes for diffraction, spectroscopy and scattering, but they interact with matter in different ways. X-rays interact largely with electrons, while neutrons interact with atomic nuclei. It follows that the magnitude of X-ray scattering depends solely on the constituent atoms electron density. Consequently, heavier atoms are stronger X-ray scatterers, as they contain more electrons. On the other hand, neutron scattering is due to nuclear and spin densities and it does not depend exclusively on atomic number, but also on the isotopic composition of the sample.

In spite of these different scattering mechanisms, however, the theories for data analysis are similar and therefore they will be described together in this section.

The first assumption that needs to be made to examine small angle scattering is that the energy and wavelength (λ) of the incident and the scattered radiations are identical; thus, we are only considering the ideal case of elastic scattering of X-rays and neutrons.

The scattering from aqueous solutions of biological macromolecules is assumed to be isotropic and only depends on the momentum transfer q, which is also called scattering vector:

$$q = \frac{(4\pi sin\theta)}{\lambda}$$

where 2θ is the scattering angle between the incident beam and the scattered radiation.

Since each atom contributes to the total scattering intensity, it is important that every measurement of the scattering of a solution of macromolecules is always accompanied by that of the corresponding solvent, in order to perform background subtraction. For both x-ray and neutron instruments, there are a variety of instrumental corrections and normalizations that are performed as part of data reduction, such as detector efficiency correction, scaling to absolute units and slit smearing.

The resulting scattering pattern $I_{\text{total}}(q)$ is given by the interference of scattered waves and can be described as:

$$I_{\text{total}}(q) = I(q) \times S(q)$$

where $I(q)$ is the form factor and $S(q)$ is the structure factor.

While the form factor $I(q)$ originates from the scattering of individual particles in solution and contains information about their structure, the structure factor $S(q)$ results from the interference of the scattered waves emitted by different particles and contains information about the structure of the solution, meaning inter-particle interactions. However, in the range of concentrations generally employed in a scattering experiment (μM, mM), we can assume that there are no inter-particle interactions. It follows that the assumption of working with an ideal (non-interacting) system can be made ($S(q)=1$), therefore only $I(q)$ is necessary to fully describe the scattering pattern.

Using small angle scattering it is possible to quickly obtain general information about the size and the shape of a macromolecule is solution, thanks to the direct computation of overall parameters from scattering data, which will be described in this section.

The initial slope of a scattering curve from a monodisperse solution of identical particles, can be approximated to a gaussian curve:

$$I(q) = I(0)\, e^{\frac{-q^2 {R_g}^2}{3}}$$

where R_g is the radius of gyration.

This approximation (Guinier's law) is valid at small q values, typically in the range $qR_g<1.3$, and allows to extrapolate the scattering intensity at zero angle $I(0)$ which cannot be determined experimentally because it is masked by the beamstop on the detector. The beamstop protects both X-ray and neutron detectors from damage or overload by the intense direct beam. R_g and $I(0)$ can therefore be calculated using the Guinier plot ($\ln[I(q)]$ versus q^2), where the slope gives R_g and the intercept with the ordinate axis gives $I(0)$.

The linearity of the Guinier region represents the first indication of sample quality. Deviations from linearity are generally signs of polydispersity or interparticle interactions.

Additional information about the macromolecule is provided by the pair-distance distribution function *p(r)*, which is computed by a Fourier transformation of the scattering data, as described by the following equation:

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 \, I(q) \frac{\sin(qr)}{qr} \, dq$$

p(r) represents the scattering data in real space and contains information on distances within the particle. For this reason, it cannot be determined directly using this equation as *I(q)* is not known over the full interval $0 \leq q \leq \infty$ and it can only be obtained with statistical and systematic errors.

However, using an indirect Fourier transform, the pair-distance distribution function can be approximated by a limited series of functions φ_i in real space. Therefore, the resulting approximated distance distribution function $p_a(r)$ can be described as:

$$p_a(r) = \sum_{i=1}^{n_s} c_i \varphi_i \, (r) \quad \text{for } 0 \leq r \leq D_{max}$$

where n_s is the number of functions φ_i and D_{max} is the maximal dimension in the particle.

Since the pair-distance distribution function contains information on distances within the particle, it follows that macromolecules with different shapes are described by different distribution functions. Figure 2.6 shows how geometrical bodies with the same maximum dimension D_{max}, but different shapes, exhibit distinct *p(r)* functions.

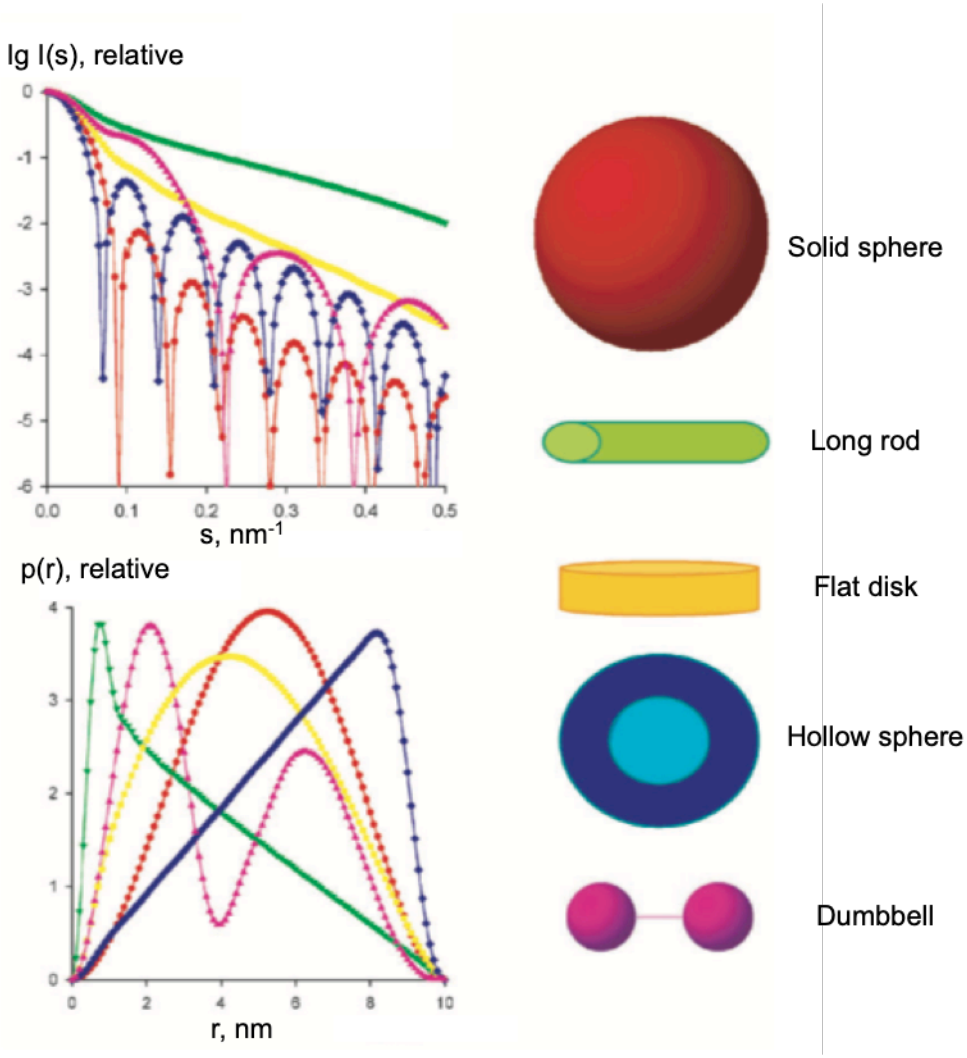


Figure 2.6: Scattering profile and pair distribution function of objects with same D_{max} but different shapes. The scattering profile and the pair distribution function of a solid sphere are described by a red line, the ones of a long rod by a green line, the ones of a flat disk by a yellow line, the ones of a hollow sphere by a blue line and the ones of a dumbbell shape representing a protein with two domains by a magenta line. (Adapted from [(Svergun et al., 2013)])

Information on the folding state of proteins, or on the compactness of the macromolecules, can be obtained directly from the scattering data using the Kratky plot ($q^2I(q)$ versus q).

As shown in Figure 2.7, the Kratky plot of a folded globular (or compact) protein shows a peak which intensity usually decays at q^{-4}. Conversely, unfolded proteins do not present any peak and instead show a plateau and a slow continuous increase at high q.

Particular attention should be taken with the interpretation of the Kratky plot of flexible systems. An extremely flexible protein will appear to have a scattering pattern similar to the one of a random coil, exhibiting a decay of the intensity as q^{-1} at high angles. However, this does not mean that the protein is present in solution as a random coil and further experiment should be performed to assess its folding state.

Qualitative analysis of the Kratky plot is also a very important step in the identification of the most appropriate tools to generate reliable low-resolution models of the macromolecule, based on its level of flexibility.

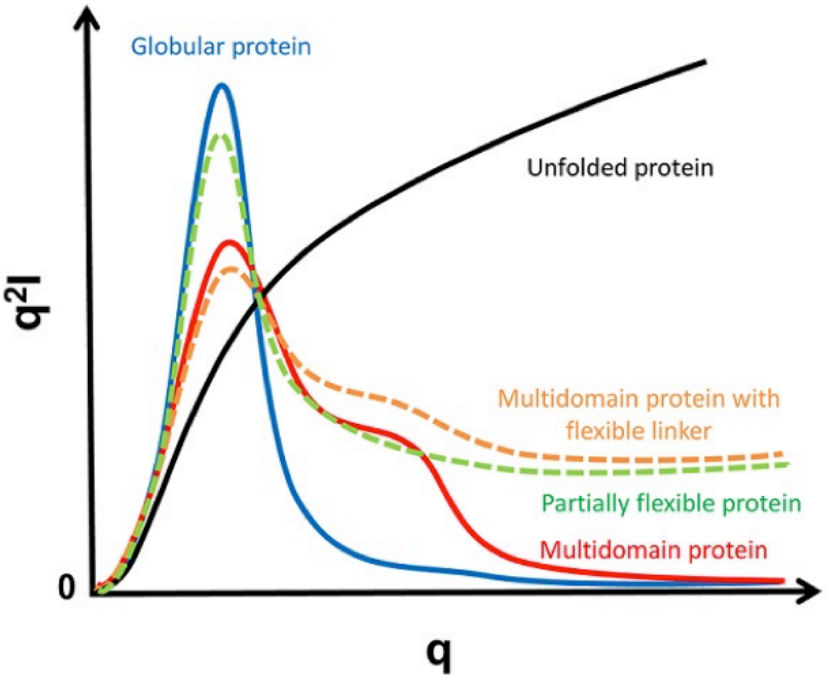


Figure 2.7: Characteristic Kratky plots of proteins exhibiting different levels of flexibility. The Kratky plot of a compact globular protein is represented by a continuous blue line, the one of a multidomain protein by a continuous red line, the one of a multidomain protein with flexible linker by a dashed orange line, the one of a partially flexible protein by a dashed green line and the one of an unfolded protein by a continuous black line. ([https://www-ssrl.slac.stanford.edu])

Three-dimensional shape determination of biological systems from their scattering pattern can be achieved by different computational analysis methods. The choice of the most appropriate modelling approach depends not only on the state of the object in solution (compact or flexible), but also on *a priori* knowledge of its association state (monomer, dimer or higher order assemblies) and on the availability of high-resolution structures for specific portions of a macromolecule or complex.

When no *a priori* structural information is available and the analysis of the Katky plot shows that the macromolecule of interest is highly compact, an *ab initio* bead modelling approach is the most straightforward method to obtain a low-resolution model of the particle.

One of the most widely used programs based on this approach is DAMMIN (Svergun). Its algorithm represents the particle as a sphere (called "search volume"), of diameter D_{max}, made of densely packed beads, where each bead in the search volume can be assigned to the solvent or to the macromolecule. Starting from a random configuration and through steps of simulated annealing, where one bead of the model is changed at each time, the algorithm calculates the theoretical scattering pattern from the generated bead model and calculates the discrepancy (χ^2) from the experimental data. The optimal configuration of densely packed beads, chosen as final model, is the one that minimise the χ^2 function (Figure 2.8).

Even though the final model generated by DAMMIN is one, *ab initio* methods do not provide a unique solution, as different models are obtained when the program is run several times on the same dataset. To reduce this intrinsic ambiguity, several utilities that allow the building of different models, and their comparison, have been developed. In particular, the workflow has been implemented by a preliminary *ab initio* bead modelling using DAMMIF, which is based on the same principles of DAMMIN but it is run several times. The models generated out of each run of DAMMIF are then compared with DAMAVER, which generates a new model with "common" features, which can be finally refined by DAMMIN.

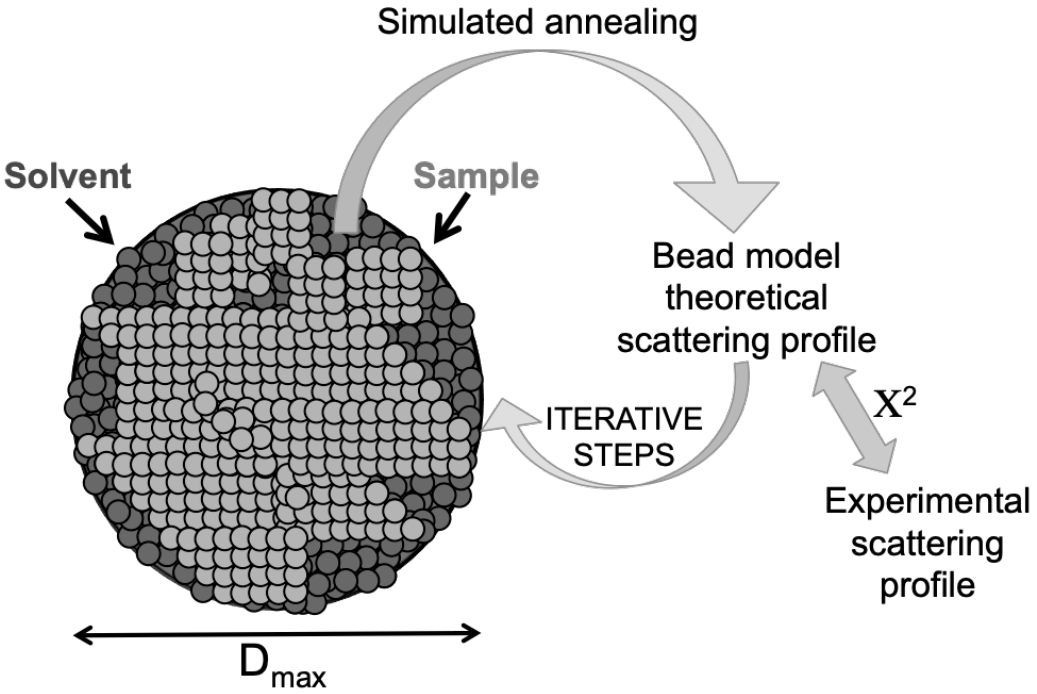


Figure 2.8: Schematic representation of *ab initio* bead modelling (DAMMIN). DAMMIN algorithm considers the particle as a sphere ("search volume"), of diameter D_{max}, made of densely packed smaller beads. Each bead in the search volume can be assigned to the solvent or to the macromolecule. One bead of the model is changed at each time through iterative steps of simulated annealing, in which the algorithm calculates the theoretical scattering pattern from the generated bead model and calculates the discrepancy (X^2) from the experimental data. The optimal configuration of densely packed beads, chosen as final model, is the one that minimise the X^2 function.

When the macromolecule of interest exhibits significant flexibility, the scattering data cannot be analysed in terms of a single beads model, as this model would represent the average of very different conformations.

Ensemble-based methods have been developed in order to account for the coexistence of multiple configurations adopted by a macromolecule in solution.

One tool based on this approach is EOM (Ensemble Optimization Method). This program also allows to use available high-resolution structures of individual subunits or domains of a protein as constraints (called “rigid bodies”) for the generation of the EOM models.

As described in Figure 2.9, this program first generates a pool of *n* independent models, based on the protein sequence and on structural information. After the creation of the pool of models, EOM runs a genetic algorithm that compares the average theoretical scattering intensity from the ensemble of *n* conformations with the experimental scattering data and selects the models that best describe the experimental data.

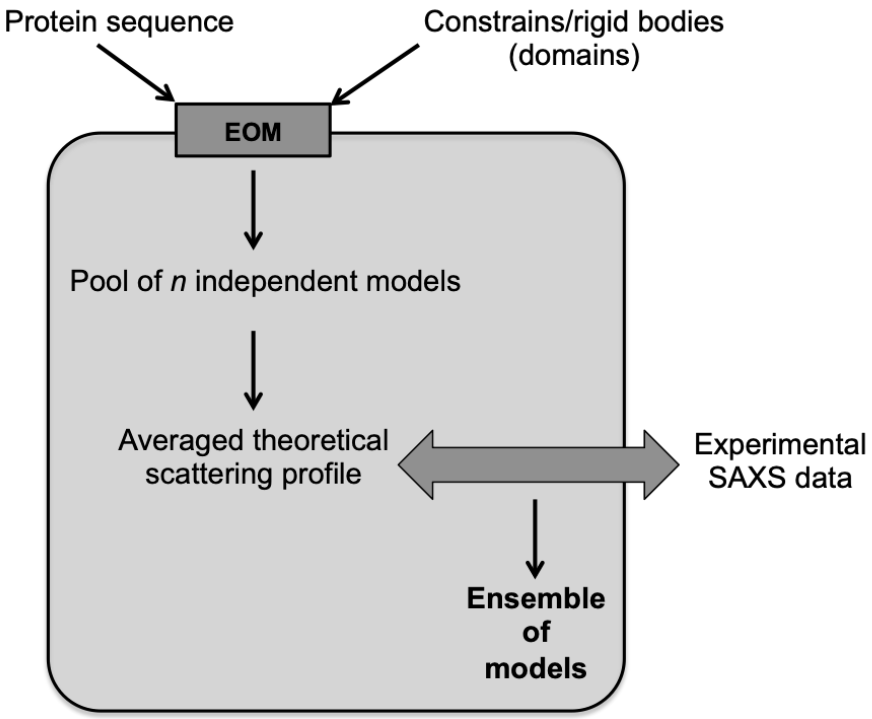


Figure 2.9: Schematic representation of ensemble optimization method (EOM). EOM generates a pool of *n* independent models, based on the protein sequence and on constraints (rigid bodies). After the creation of the pool, EOM runs a genetic algorithm that compares the average theoretical scattering intensity with the experimental data and selects the models that best describe the experimental scattering profile.

2.6 Transmission electron microscopy

Transmission electron microscopy (TEM) is a powerful tool that allows to determine the three-dimensional structure of biological macromolecules in native-like conditions. Recent innovations in TEM instrumentations, in particular the development of direct electron detectors and phase plates, and improvement of image processing methods, led to a “resolution revolution” with the establishment of electron microscopy as a tool to obtain near-atomic resolution structures, previously achieved only by X-ray crystallography or NMR spectroscopy.

Figure 2.11 shows the main components of an electron microscope. Electrons are generated by an electron gun (cathode) and then accelerated by an anode. The electron beam is focused onto the specimen by a condenser system and the image is subsequently produced and magnified by a set of electrostatic and electromagnetic lenses. The electron image is then transformed into something perceptible by the human eye and then recorded by a digital camera or by a direct electron detector.

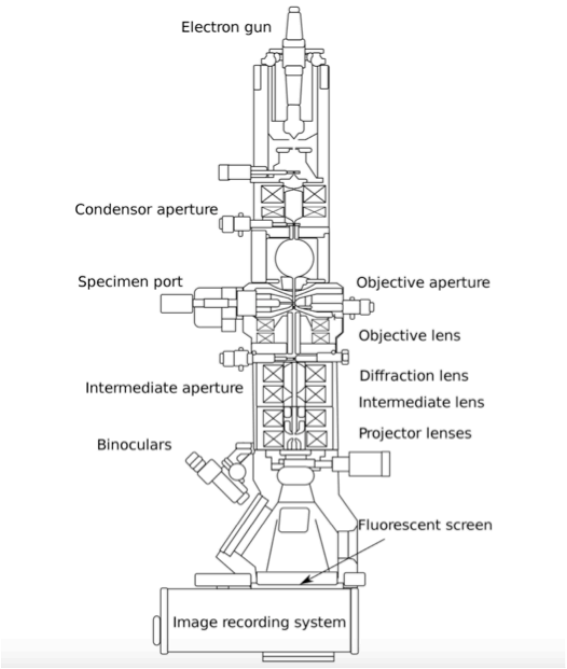


Figure 2.10: Schematic representation of TEM microscope. The beam of electrons is generated by an electron gun and it is focused into a small coherent beam by the use of lenses. The beam hits the specimen and its transmission generates the image, which is then recorded and used for image processing.

Since electrons interact strongly with matter, electron microscopy is performed under vacuum conditions and for this reason the sample must be in the solid state for imaging.

There are two approaches that allow to achieve this. The first method is called “negative staining” and it consists in embedding the sample in a thin heavy metal salt layer, traditionally uranyl acetate, which allows to visualize the envelope of the macromolecule of interest by exclusion (Figure 2.12). The protein suspension is firstly deposited on an EM copper grid covered with a 5-10 nm carbon film and then incubated with 2% uranyl acetate solution, which is subsequently blotted and allowed to dry. This method is quick and simple, and it allows to obtain a low-resolution model of the macromolecule of interest, for this reason it is usually adopted as first step in electron microscopy studies. However, this technique presents some difficulties because some protein complexes cannot withstand the drying process and they collapse. Also, the models obtained from negative staining electron microscopy might not reflect the real conformation of the protein in solution because of the distortion induced by the metal layer.

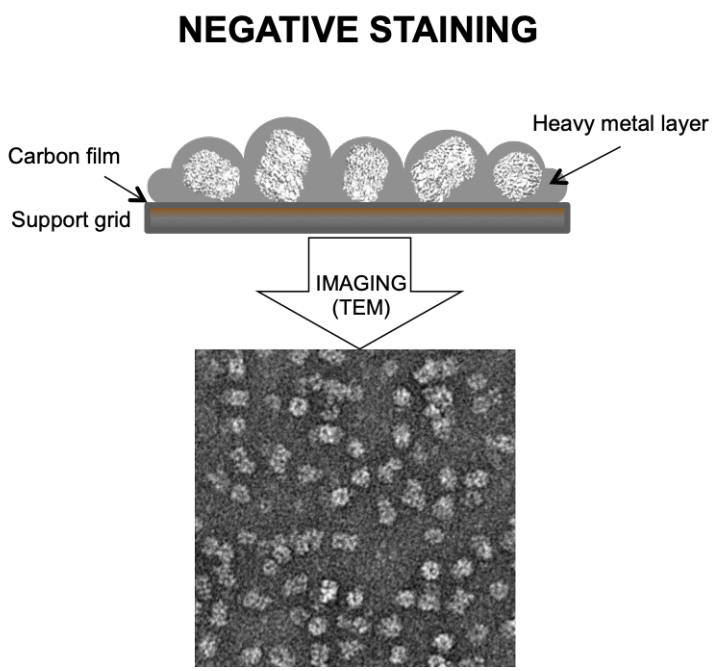


Figure 2.11: Schematic representation of negative staining principle. The protein suspension is deposited on an EM copper grid covered with a 5-10 nm carbon film, let sit for 2 minutes and then incubated with 2% uranyl acetate solution for 5 minutes, which is subsequently blotted and allowed to dry.

A second method, cryo-electron microscopy (cryo-EM), has been developed in order to overcome these limitations (Figure 2.13). Cryo-EM allows to obtain more accurate high-resolution models. This method consists in applying the protein suspension on a holey (Quantifoil) copper or gold EM grid, which can be also covered by a very thin layer of carbon (1-2 nm) or graphene oxide to allow better sample distribution. The sample is then blotted, allowing the macromolecules to enter the grid holes, and subsequently plunge-frozen in liquid ethane, which is kept cold by liquid nitrogen. Fast freezing allows to bring the sample to the solid state in native-like conditions, where its hydration state is maintained in vitrified ice, and prevents the formation of ice crystals, which would be detrimental to sample quality. Moreover, the specimen needs to be maintained at low temperature throughout the whole process, from sample preparation to imaging, which reduces the effect of radiation damage. Contrary to negative staining, cryo-EM is a time-consuming technique which requires extensive screening of different sample preparation conditions in order to optimise particles distribution and ice thickness.

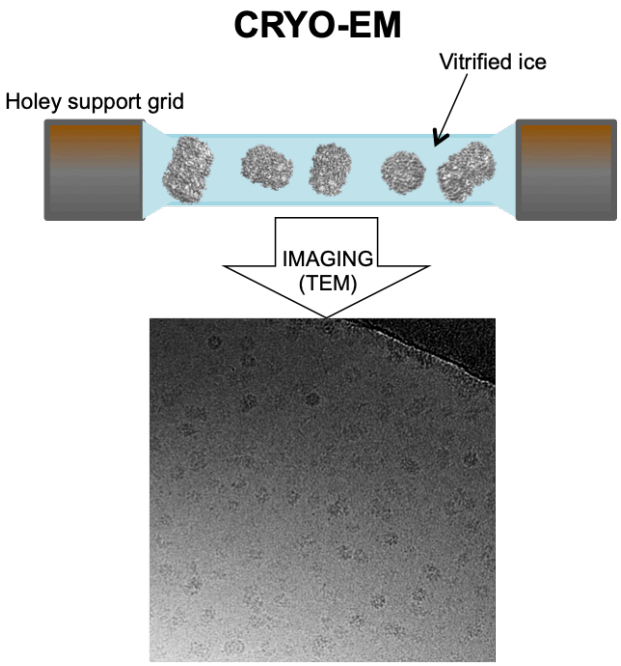


Figure 2.12: Schematic representation of cryo-EM principle. The protein suspension is deposited on a holey (Quantifoil) copper or gold EM grid (with or without a thin (1-2 nm) layer of carbon or graphene oxide. The sample is then blotted and subsequently plunge-frozen in liquid ethane, which is kept cold by liquid nitrogen.

Single-particle analysis (SPA) coupled to cryo-EM allows to achieve high-resolution structural information of individual proteins or complexes. A substantial part of a single-particle project consists on processing the acquired images. To simplify data processing, several softwares have been developed for SPA, such as EMAN2 and RELION. Even though these packages are based on different approaches and algorithms, they share the same basic concepts. The typical SPA workflow is illustrated in Figure 2.14.

The development of direct electron detectors allows to record movies in which the electron dose is fractionated into a series of frames, which can be then aligned and averaged. This allows not only to account for initial stage drifting and beam-induced movements and correct it by excluding the first frames, but also to optimise the signal-to-noise ratio by excluding the last frames that will have a lower quality due to radiation damage.

Moreover, CTF estimation allows to further evaluate image quality. CTF parameters depend on instrument features (voltage and spherical aberration) and microscopy settings used during the data collection (astigmatism, defocus and amplitude contrast). The subsequent correction for these parameters allows to additionally improve the quality of the dataset by selecting the micrographs which contain high-resolution information.

In single-particle analysis, individual reprojections of the object in different orientations are picked and then extracted from the micrographs. Particles are then grouped in 2D class averages by iterative alignments, which allows to divide the dataset in homogeneous subsets with increased signal-to-noise ratio.

A limited number of 2D class averages, representing different particle orientations, is then used to obtain an initial 3D model where angles and relationships between the reprojections are calculated through computational methods. The resulting model that best describe the object is then used as low-resolution template for subsequent 3D refinement.

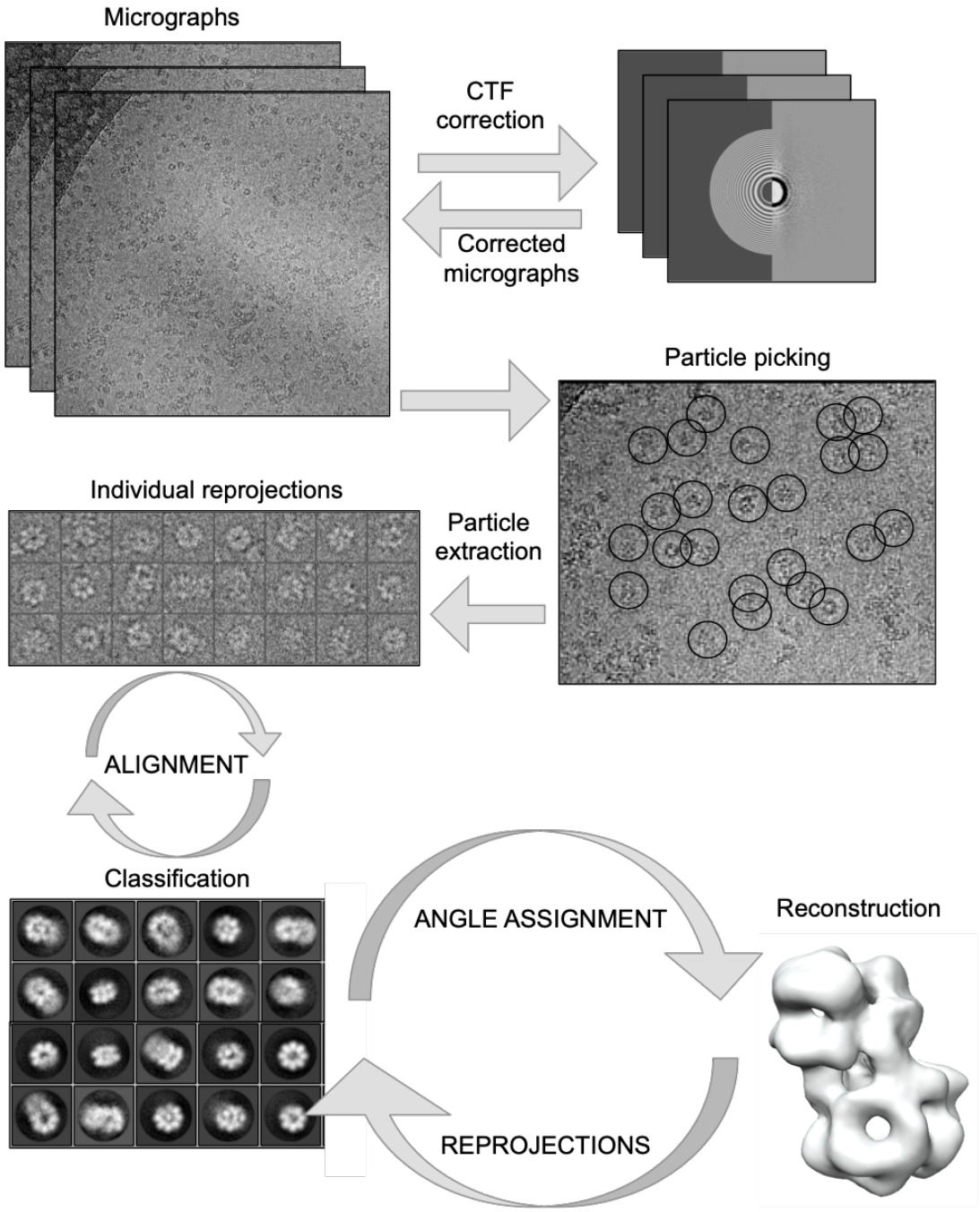


Figure 2.13: Schematic representation of single particle analysis. Individual particles are picked from micrographs, which have been previously aligned and CTF corrected. Particles are then extracted to obtain individual reprojections that are successively subjected to several rounds of alignments in order to generate 2D class averages. These are then used to obtain an initial 3D model, for which angles and relationships between the reprojections are calculated through computational methods. The resulting model that best describe the object is then used as low-resolution template for subsequent iterative 3D refinement.

The improvement of the refinement can be monitored by evaluation of the Fourier Shell Correlation (FSC) curve, which provides information on the signal-to-noise ratio as a function of spatial frequency and resolution of the 3D density map. However, the “resolution” of cryo-EM maps cannot be univocally defined as several criteria are used for its determination. For example, the resolution can be defined as the spatial frequency at which the FSC curve is 0.5 or 0.143, the latter threshold being selected based on correlation studies between results obtained in EM and in x-ray crystallography.

The interpretation of low/intermediate resolution EM maps is highly enhanced by the availability of atomic resolution structures of some portions of the protein or some components of a complex. Indeed, structures from NMR spectroscopy or X-ray crystallography studies, can be fitted in EM density map to validate the model and provide detailed information on feature localisation. On the other hand, caution should be taken with the interpretation of low-resolution EM maps. In fact, they only reveal the overall architecture of a protein (or complex) and no additional information, other than domains arrangement or components localisation, can be extrapolated by fitting high-resolution structures into the density map.

Chapter 3

Human Staufen protein

3.1 Introduction

Staufen (Stau) is a dsRNA binding protein, which was originally identified in *Drosophila melanogaster*, where it plays an essential role in oocyte development (St Johnston et al., 1991, St Johnston and Nusslein-Volhard, 1992). It is well conserved from nematodes to humans and, depending on the species, it is composed of four or five dsRNA-binding domains (RBDs) (Wickham et al., 1999b). In humans, there are two Staufen paralogs: Stau1 and Stau2. The hStau1 gene encodes two main isoforms (55 and 63 kDa) and an additional isoform with six residues inserted into hStau1 RBD3 (Duchaine et al., 2000), whereas the hStau2 gene encodes at least four isoforms (52, 56, 59 and 62 kDa): all these isoforms are generated by alternative pre-mRNA splicing and/or polyadenylation (Park and Maquat, 2013). Much of our knowledge on human Stau1 is based on the study of its hStau1^{55} isoform. The hStau1^{55} isoform is associated with 40S and 60S ribosomal subunits and co-localises with the rough endoplasmic reticulum (Luo et al., 2002, Wickham et al., 1999b, Marion et al., 1999). Besides the biochemical studies on its ribosome association, hStau1^{55} has been characterised biochemically in the context of mRNA decay (Kim et al., 2005) and cell cycle control (Boulay et al., 2014). Whilst Stau2 is expressed only in the neuromuscular system and it is mostly involved in mRNA transport at particular sites of the post-synaptic muscles, Stau1 is ubiquitously expressed (Bélanger et al., 2003, Lebeau et al., 2008, Vessey et al., 2008, Ravel-Chapuis et al., 2012, Peredo et al., 2014). Even though Stau1 and Stau2 exhibit different tissue localization, they have been shown to be involved in the same mechanisms of RNA regulation, such as mRNA transport (Martel et al., 2010, Martel et al., 2006, Ramasamy et al., 2006, Vessey et al., 2008, Ravel-Chapuis et al., 2012, Ferrandon et al., 1994), Staufen-mediated mRNA decay (Cho et al., 2013, Cho et al., 2012, Kim et al., 2014, Park and

Maquat, 2013, Gong et al., 2009, Gong and Maquat, 2011, Gong et al., 2013, Kim et al., 2005, Park, 2013) and regulation of protein translation (Ravel-Chapuis et al., 2012, Elbarbary et al., 2013, Bonnet-Magnaval et al., 2016). Among the several cellular processes in which Stau protein is involved we can include myogenic differentiation (Ravel-Chapuis et al., 2014), stress granule formation (Ravel-Chapuis et al., 2016), regulation of adipogenesis (Cho et al., 2013, Cho et al., 2012), progression of the cell cycle (Boulay et al., 2014) and cellular differentiation (Kretz, 2013, Gautrey et al., 2005, Gautrey et al., 2008, Peredo et al., 2014). Stau1 and Stau2 are also central players in virology: they play a role in HIV infection by favoring vRNA encapsidation (Banerjee et al., 2014, Chatel-Chaix et al., 2008, Chatel-Chaix et al., 2004, Mouland et al., 2000), in hepatitis C infection by transporting viral RNA to the site of translation and replication in the cells (Blackham and McGarvey, 2013, Dixit et al., 2016) and they are also required for efficient multiplication of influenza A virus (de Lucas et al., 2010). The variety of cellular processes in which Staufen is implicated suggests that this protein might adopt different binding modes with its diverse RNA targets and that structurally distinct RNA-Staufen complexes mediate the recruitment of effectors *via* protein-protein and/or protein-RNA interactions.

The interaction between Stau1 and its RNA substrates has been initially characterised using molecular and cellular biology tools on hStau1^{55}. The evidence that multiple copies of hStau1^{55} can bind a single dsRNA is based on studies showing that two hStau1^{55} molecules bind the hARF1 RNA in cells, which is a target of Staufen-mediated mRNA decay (SMD) (Martel et al., 2010) and mRNA containing as many as 250 CUG repeats that typify patients with myotonic dystrophy *in vitro* (Ravel-Chapuis et al., 2012). Furthermore, hStau1^{55} stabilizes relatively large imperfect base-pairing regions formed between mRNAs and long noncoding RNAs (Gong and Maquat, 2011). Additionally, multiple hStau1 molecules bind in tandem to the same dsRNA to efficiently recruit the ATP-dependent helicase hUPF1. Genome-wide analysis (Furic et al., 2008, Laver et al., 2013) and hiCLIP (RNA hybrid and individual-nucleotide resolution ultraviolet cross-linking and immunoprecipitation) (Fernandez Moya and Kiebler, 2015) of Staufen-associated mRNAs identified secondary structures that confer target specificity (Ricci et al., 2014). The hiCLIP experiments confirmed Arf1 RNA as a substrate for Stau1, adding evidence for the presence of a sbsRNA on the 3' UTR

of the X-box binding protein 1 (XBP1) mRNA (Fernandez Moya and Kiebler, 2015). However, the molecular mechanism of Staufen/RNA recognition is still unclear (de Lucas et al., 2014), since even the recent structure of the RBD3-RBD4 in complex with dsRNA does not provide a conclusive answer to whether or not this is sequence specific (Lazzaretti et al., 2018).

Staufen is characterized by two conserved consensus amino acid motifs that fold into functional double-stranded RNA binding domains (RBD3, RBD4); hStau1 contains two other RBDs (RBD2, RBD5) that are unable to bind RNA and, relative to hStau1, hStau2 has an additional RBD1 and only a partial RBD5 (Allison et al., 2004, Wickham et al., 1999a, Furic et al., 2008, Buchner et al., 1999, Duchaine et al., 2002). The tubulin-binding domain (TBDs), which is involved in mRNA transport on the cytoskeleton, shares only 18% identity between hStau1 and hStau2 (Park and Maquat, 2013). The functional activation of a number of dsRNA-binding proteins requires that they self-associate or associate with other dsRNA-binding proteins (Park et al., 2013). A Staufen swapping motif (SSM) has been identified between the TBD and RBD5. The SSM is necessary for the homodimeric or heterodimeric interactions between Stau1 and Stau2. This dimerization is indispensable for Staufen activity during Staufen-mediated mRNA decay (Park, 2013, Martel et al., 2010, Gleghorn et al., 2013). RBD5 was also identified as the major determinant for protein–protein interaction *in vivo*, contributing to ~50% of the Stau1 self-interaction. A recent SEC-MALLS report on purified protein also shows that, in the absence of RNA, SSM/RBD5 promotes dimerization (Lazzaretti et al., 2018). The importance of RBD2 in dimerization is less clear. BRET assays aimed at the study of hStau1^{55} multimerization show that RBD2 (amino acids 37–79, isoform hStau1^{55}) interacts with full-length hStau1 (Martel et al., 2010). On the other hand, recombinant purified RBD2-RBD3 suggests that the contribution of RBD2 to hStau1^{55} dimerization, while existing, is relatively minor (Martel et al., 2010, Park et al., 2013).

To date, the analysis of the three-dimensional structure of Stau proteins has focused on studies of truncated versions of the protein in isolation or in complex with either short RNA sequences or truncated versions of interacting proteins (Figure 3.1). The NMR structure of *Drosophila* RBD3 in isolation first confirmed that this construct is organised in the typical α-β-β-β-α fold (PDB ID: 1STU)

(Bycroft et al., 1995). Mouse Stau2 RBD4 in the absence of dsRNA also showed the α-β-β-β-α fold (PDB ID: 1UHZ). The structure of human Stau1 SSM/RBD5 solved by X-ray crystallography revealed a domain swapped dimer, which is responsible for mediating hStau1 dimerization (PDB ID: 4DKK) (Gleghorn et al., 2013). The X-ray crystal structure of the complex between RBD5 and Miranda, a multi-domain scaffold protein involved in *Drosophila* neuroblast asymmetric division, showed two RBD5 domains symmetrically bound to the Miranda dimeric coiled coil region through their exposed β-sheet faces, revealing a previously unrecognized protein interaction mode for dsRBDs (PDB ID: 5CFF) (Jia et al., 2015). The solution structure of *Drosophila melanogaster* Stau RBD3 bound to a 12-bp stem-loop RNA, determined by NMR spectroscopy, revealed the interaction of the canonical α-β-β-β-α dsRBD fold with dsRNA (PDB ID: 1EKZ) (Ramos et al., 2000b, Ramos et al., 2000a). The crystallographic structure of the RBD3-RBD4 construct, bound to dsRNA as a dimer (monomers A and B), shows that the interaction surface with the RNA spans the major groove and the two adjacent minor groove surfaces. Furthermore, RBD3B is bound on the opposite side of the RNA molecule, in an antiparallel orientation to RBD3A, whereas density for the second RBD4 is missing (Lazzaretti et al., 2018).

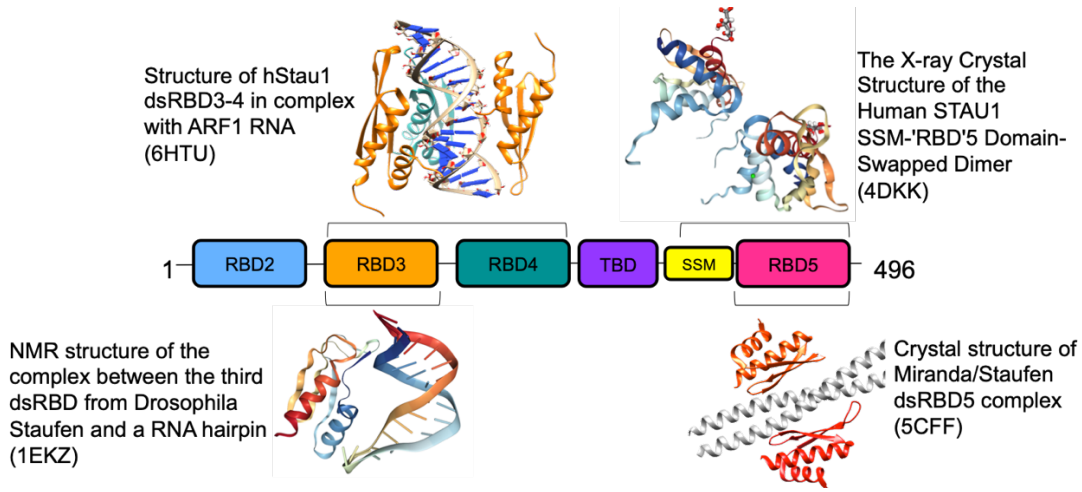


Figure 3.1: Domains organization of hStaufen1 (isoform 55 kDa) and available structural information. hStaufen1[55] domains are shown as coloured rectangles (RBD2: blue, RBD3: orange, RBD4: teal, TBD: purple, SSM: yellow, RBD5: pink). Available structures of individual and tandem domains, in complex with RNA or polypeptide molecules, are represented in ribbon diagrams and described in the figure.

Human, *Drosophila*, and *C. elegans* Stau bind dsRNA without apparent sequence specificity in vitro (St Johnston et al., 1992, Marion et al., 1999, Wickham et al., 1999a, Ramos et al., 2000b, LeGendre et al., 2013, Wang et al., 2015). Yet, in all these systems, Stau proteins associate with specific RNA targets (St Johnston and Nusslein-Volhard, 1992, Ferrandon et al., 1994, Li et al., Mallardo et al., 2003, Kim et al., 2005, Kim et al., 2007, Furic et al., 2008, Heraud-Farlow et al., 2013, Laver et al., 2013, LeGendre et al., 2013, de Lucas et al., 2014, Ricci et al., 2014, Sugimoto et al., 2015), challenging the hypothesis that the binding of a dsRBDs to dsRNA has generally been considered to be dependent on the target RNA structure, rather than sequence. On the other hand, Bono and co-workers recently showed that, in addition to the interactions with the sugar–phosphate backbone previously shown for RBD3 (Ramos et al., 2000c), both domains of hStau1 directly contact RNA bases in the minor groove of ARF1 SBS. Indeed they also show that specific base recognition is relevant *in vivo* and may therefore contribute to the overall sequence selectivity by Stau, possibly together with additional regions of the protein or with other regulators (Lazzaretti et al., 2018). The macromolecular interaction events that happen downstream of Staufen RNA binding are still unknown.

We used an integrated structural biology approach, combining homology modelling, small angle X-ray scattering, NMR, hydrodynamic methods and cryo-EM, to characterize the structure of the human full-length Staufen1 protein to better understand the structural basis for its functional plasticity. More specifically, we aim to decipher the structural components that play a role in the ability of this protein to interact *in vivo* with a wide variety of RNA substrates and protein interactors in order to mediate many diverse functions in RNA metabolism. To interpret this structure, we also studied its truncation mutant hStau1^{55}_ΔRBD2, its individual domains and several tandem domain fragments. We confirmed that RBD2 has an important role in protein oligomerization, as well as reporting for the first time its effect on protein solubility. Our data show for the first time that hStau1^{55} adopts an elongated conformation in solution. Furthermore, in the absence of RNA, RBD3 and RBD4 are connected by a linker that is very flexible in solution, and they do not interact with one another.

3.2 Materials and Methods

3.2.1 Staufen proteins

3.2.1.1 Cloning of individual and tandem domains

Individual and tandem hStau1[55] domains were amplified by Polymerase Chain Reaction (PCR) from pRSET-B-Stau1[55] vector (Kim et al., 2005) as described in Table 3.1. Different annealing temperatures were used for primers corresponding to different constructs (shown as "*" in Table 3.1). Primers sequences and constructs boundaries are indicated in Table 3.2. PCR products for RBD2, RBD2-RBD3, RBD3, RBD3-RBD4, RBD4 and TBD were purified using the QIAquick PCR Purification Kit (QIAGEN). PCR products for RBD4-TBD, SSM/RBD5 and TBD-SSM/RBD5 were purified from 2% agarose gel run in TBE using the MinElute Gel Extraction Kit (QIAGEN). Purified PCR products and pET28-a vector were digested with NdeI and HindIIIHF (NEB) for 3 hours at 37°C. Digested vector was purified from 1% agarose gel run in TBE using the MinElute Gel Extraction Kit (QIAGEN), whereas digested inserts were purified using the QIAquick PCR Purification Kit (QIAGEN). Ligations between the vector and the individual inserts (in ratio 1:3) were performed using the Quick Ligation Kit (NEB) for 5 minutes at RT. *E. coli* XL1-Blue cells were transformed by heat-shock method with 2.5 µl of the ligation reactions and plated in LB agar plates containing 50 µg/ml kanamycin. After overnight, for each of the transformations, 15 ml of LB, supplemented with 50 µg/ml kanamycin, were inoculated with a single colony and the cultures were grown at 37°C o/n. Subsequently, plasmid DNA was purified using the Wizard Plus SV Minipreps DNA Purification Systems and sequenced.

PCR step	Temperature (°C)	Time (sec)	
Initial denaturation	98	30	
Denaturation	98	10	40 cycles
Primers annealing	*	20	
Extension	72	30	
Final extension	72	120	
Hold	10	hold	

Table 3.1: PCR program used to amplify individual and tandem domains. Primers annealing temperature is indicated in the table by "*" and it varies as following for the constructs: RBD2 and RBD2-RBD3 (62°C), RBD3 and RBD3-RBD4 (60°C), RBD4 (65°C), SSM/RBD5, RBD4-TBD and TBD-SSM/RBD5 (67°C), TBD (70°C).

Construct	Boundaries (aa)	Oligos
RBD2	1-85	fw: 5'-CCCCGAACATATGATGAAACTTGGAAAAAAACCAATGTATAA-3' rev: 5'-TGCTTAAAGCTTTCATCATCATGGCAGGGGCTCATTCTGCA-3'
RBD3	95-172	fw: 5'-CCCCGAACATATGTCCGAAGAAGAAAATCTCAATAAAT-3' rev: 5'-TGCTTAAAGCTTTCATCATCACGGTAACTTCTTCAGCTCCT-3'
RBD4	198-275	fw: 5'-CCCCGAACATATGCCAGAATATGGCCAGGGGAT-3' rev: 5'-TGCTTAAAGCTTTCATCATCAGACTTTGAAACCAAGGATCTCCA-3'
TBD	283-366	fw: 5'-CCCCGAACATATGCCCGCACTCAAGTCAGAGGA-3' rev: 5'-TGCTTAAAGCTTTCATCATCAGGCAGGATTCGGAGCTGCC-3'
SSM/RBD5	363-476	fw: 5'-CCCCGAACATATGCCGAATCCTGCCAAGGCCAC-3' rev: 5'-TGCTTAAAGCTTTCATCATCAGTCCAACTCAGACAGCAACTTTAAG-3'
RBD2-RBD3	1-172	fw: 5'-CCCCGAACATATGATGAAACTTGGAAAAAAACCAATGTATAA-3' rev: 5'-TGCTTAAAGCTTTCATCATCACGGTAACTTCTTCAGCTCCT-3'
RBD3-RBD4	95-275	fw: 5'-CCCCGAACATATGTCCGAAGAAGAAAATCTCAATAAAT-3' rev: 5'-TGCTTAAAGCTTTCATCATCAGACTTTGAAACCAAGGATCTCCA-3'
RBD4-TBD	198-366	fw: 5'-CCCCGAACATATGCCAGAATATGGCCAGGGGAT-3' rev: 5'-TGCTTAAAGCTTTCATCATCAGGCAGGATTCGGAGCTGCC-3'
TBD-SSM/RBD5	283-476	fw: 5'-CCCCGAACATATGCCCGCACTCAAGTCAGAGGA-3' rev: 5'-TGCTTAAAGCTTTCATCATCAGTCCAACTCAGACAGCAACTTTAAG-3'

Table 3.2: Primers used for the cloning of individual and tandem domains from pRSET-B-Stau1[55] vector.

3.2.1.2 Proteins overexpression and purification for SAXS, AUC and EM experiments

Recombinant proteins were overexpressed in *E. coli* Rosetta™(DE3) pLysS cells transformed by heat-shock with pRSET-B vectors containing hStau1^{55}_FL (Kim et al., 2005) and hStau1^{55}_ΔRBD2 or with pET28a containing individual domains (RBD2, RBD3, RBD4, TBD, SSM/RBD5) and tandem domains (RBD2-RBD3, RBD3-RBD4, RBD4-TBD, TBD-SSM/RBD5). Starting cultures were grown in LB medium containing 50 µg/ml ampicillin (for hStau1^{55}_FL and hStau1^{55}_ΔRBD2) or 50 µg/ml kanamycin (for individual and tandem domains) and 34 µg/ml chloramphenicol at 37°C o/n. 1 ml of overnight culture was inoculated in 1 L of TB medium supplemented with antibiotics and cells were grown to an OD_{600}=0.6. Protein overexpression was induced by adding 0.5 mM IPTG and culturing the cells at 27°C overnight. Cells were harvested by centrifugation at 5000 rpm on a Beckman Avanti™ J-20 XP centrifuge with JLA 8.1000 rotor for 20 minutes and washed once with PBS; cell pellets were aliquoted and stored at -80°C. Frozen aliquots were thawed and lysed by sonication in 20 ml lysis buffer (25 mM HEPES pH 7.5, 1 M GndCl, 20 mM imidazole, 1% Triton X-100, 400 µl Complete EDTA-free protease inhibitors (50X) and 2 µl benzonase) followed by 30 min of incubation on ice. Soluble protein extracts were separated from cell pellets by centrifugation at 15000 rpm for 30 min. 6-His tagged Staufen1 proteins were purified by nickel chromatography using HisTrap FF columns (GE Healthcare) equilibrated in washing buffer (25 mM HEPES pH 7.5, 1 M GndCl, 20 mM imidazole). Elution was performed with a 20 mM-1 M imidazole gradient. Fractions containing the protein of interest were pooled and concentrated to 10 mg/ml prior to size exclusion chromatography on a Superdex 200 Increase 10/300 GL (hStau1^{55}_FL and hStau1^{55}_ΔRBD2) or a Superdex 75 Increase 10/300 GL (individual and tandem domains) gel filtration column (GE Healthcare) in buffer A (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 200 mM L-Arg HCl). Overexpression and purification efficiency were monitored by SDS-PAGE analysis on 12% or 18% gels stained with SimplyBlue SafeStain (Invitrogen). Fractions were also analyzed by Western Blot: the membrane was incubated o/n at 4°C with anti-His antibody conjugated with alkaline phosphatase (1:4000) and developed using the SIGMA-FAST™ BCIP/NBT reagent.

3.2.1.3 Proteins overexpression and purification for NMR analysis

E. coli Rosetta™(DE3) pLysS cells were transformed by heat-shock with vectors carrying inserts for hStau1^{55}_FL, hStau1^{55}_ΔRBD2, individual domains (RBD2, RBD3, RBD4, TBD, SSM/RBD5) and tandem domains (RBD2-RBD3, RBD3-RBD4, RBD4-TBD, TBD-SSM/RBD5). Starting cultures were grown as described above and the following day cells were cultured in TB medium supplemented with antibiotics to OD_{600}=2. Cells were harvested by centrifugation at 5000 rpm on a Beckman Avanti™ J-20 XP centrifuge with JLA 8.1000 rotor for 20 minutes, resuspended in M9 minimal medium (1x M9, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) and grown for 30 min at 37°C. Successively, the minimal medium was supplemented with 400 µl Thiamine [50 mg/ml], filter-sterilized glucose [3 g/L] and filter-sterilized ISOGRO®-15N Powder-Growth Medium [1 g/L]. Protein overexpression was induced by adding 0.5 mM IPTG and culturing the cells at 27°C overnight. Purification of Staufen proteins was performed as described above, but replacing HEPES with 20 mM potassium phosphate buffer pH 7.5 in all buffers used. To prevent formation of disulfide bonds, 2 mM TCEP was added to the purified proteins.

3.2.1.4 Analytical ultracentrifugation (AUC)

Sedimentation velocity experiments were performed at 50,000 rpm, using a Beckman Optima analytical ultracentrifuge equipped with an An-50Ti rotor. Data were recorded using both the absorbance (at 280 nm and 260 nm) and interference optical detection systems. The density and viscosity of the buffer were measured experimentally using a DMA 5000M densitometer equipped with a Lovis 2000ME viscometer module. The partial specific volume for the protein was calculated using SEDNTERP from the amino acid sequence. Data were processed using SEDFIT, fitting to the c(s) model. Figures were made using GUSSI (Lebowitz et al., 2002).

3.2.1.5 SEC-MALLS

SEC-MALLS experiments were performed using a Superdex 200 10/300 Increase column (GE Healthcare) connected to an AktaPure 25 System (GE Healthcare). The protein sample (100 µL) was loaded onto the gel filtration column and eluted with one column volume (24 mL) of buffer A, at a flow rate of 0.7 ml/min. The eluting protein was monitored using a DAWN HELEOS-II 18-angle light scattering detector (Wyatt Technologies) equipped with a WyattQELS dynamic light scattering module, a U9-M UV/Vis detector (GE Healthcare), and an Optilab T-rEX refractive index monitor (Wyatt Technologies). Data were analysed by using Astra (Wyatt Technologies) using a refractive increment value of 0.185 mL/g.

3.2.1.6 Small Angle X-ray Scattering (SAXS) and Modelling

SAXS data for hStau1^{55}_FL and hStau1^{55}_ΔRBD2 were collected on a Nano-inXider instrument (1x10^{6} ph/s) (Xenocs SA, Grenoble, France), at the ISIS Neutron and Muon Source (Harwell, UK). 80 µl of sample (hStau1^{55}_FL at 40 µM and hStau1^{55}_ΔRBD2 at 60 µM), were loaded using the Xenocs “low noise” flow cell chamber connected to an auto-sampler unit, and data were acquired in high resolution mode every 10 minutes for 120 seconds, for a total of 2 hours. SAXS data for Staufen1 proteins were reduced to 1D curves and buffer-subtracted using the Foxtrot package and then input to EOM 2.0 (Tria, 2015) (ATSAS suite). High-resolution structures of individual domains were used as rigid bodies and constraints in the model generation. In our analysis, the modelled structures of the individual domains of Staufen1 were obtained using the Phyre2 web portal (Kelley et al., 2015). Human Staufen1 RBD3 was modelled by homology based on the NMR structure of Drosophila melanogaster Staufen RBD3 (PDB ID: 1EKZ) (Ramos et al., 2000b, Ramos et al., 2000a). The structure of human Staufen1 RBD4 was obtained by homology modelling based on the mouse RBD4 (PDB ID: 1UHZ). The structures of SSM and RBD5, were extracted from the structure solved by X-ray crystallography (PDB ID: 4DKK) (Gleghorn et al., 2013) and treated as two separate domains in this analysis, allowing complete inter-domain loop flexibility. To obtain a more complete set of structural information to use as

constraints for the interpretation of the SAXS data, the sequences of the 6-His+linker+RBD2 domain of hStau1^{55}_FL, the 6-His+linker of hStau1^{55}_ΔRBD2 and the TBD were modelled using the Phyre2 server (Kelley et al., 2015). The program EOM 2.0 (Tria, 2015) was used to obtain the SAXS models of hStau1^{55}_FL protein and deletion mutants. A pool of 10,000 independent models was generated, based on the sequence of hStau1^{55}_FL, or hStau1^{55}_ΔRBD2, and on constraints we generated by homology modelling.

SAXS data for individual domains and tandem domains were collected at B21, Diamond Light Source (Harwell, UK). 55 µl of each protein sample (~10 mg/ml) was loaded onto a Superdex 75 Increase 10/300 GL column (GE Healthcare), controlled by an Agilent HPLC system, coupled to an in-vacuum SAXS flow cell. HPLC-SAXS traces were processed using ScÅtter. Data were analysed using different strategies depending on their flexibility level. Models for individual domains RBD3 and RBD4 were obtained using ScÅtter and DAMMIN and the tandem domain TBD-SSM/RBD5 was modelled using BUNCH (ATSAS). All the other domains, showing higher degree of flexibility, were modelled with EOM using homology models obtained from Phyre2, as described above.

3.2.1.7 NMR spectroscopy

hStau1^{55}_FL, hStau1^{55}_ΔRBD2, individual and tandem domains were studied by NMR spectroscopy. 30 µl of D_2O were added to 570 µl of protein in 20 mM potassium phosphate pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 200 mM L-Arg.HCl, 2 mM TCEP at a suitable concentration for NMR experiments (Table 3.3). 2D ^{15}N TROSY experiments (trosyf3gpphsi19; Czisch et al., 1998; Pervushin et al.,1998) were recorded at 298 K on a Bruker AVANCE IIIHD spectrometer operating at 14.1 T equipped with a 5 mm TCI cryoprobe and running TopSpin software. Spectra of hStau1^{55}_FL and hStau1^{55}_ΔRBD2 were recorded with 1024 and 128 complex points and sweep widths of 10775.862 Hz and 1823.747 Hz in ^{1}H and ^{15}N, respectively. The data were processed by applying Lorentz to Gauss (GM; LB=-7.72; GB=0.1667) window function and zero filled to 2048 complex points before Fourier transformation in the ^{1}H dimension. In the ^{15}N dimension, linear prediction with 48 coefficients was applied extending the data to 512

complex points. This was followed by a sinebell squared (QSINE; SSB=2.25) window function and Fourier transformation. Spectra of the shorter Staufen fragments were acquired and processed similarly, except that only 78 complex points were recorded in the ^{15}N dimension. Data were processed using the Bruker TopSpin software and figures were generated using CCPN analysis 2.4 (Vranken et al., 2005).

Construct	Concentration for NMR acquisition (µM)
hStau1^{55}_FL	66
hStau1^{55}_ΔRBD2	400
RBD2	33.3
RBD3	128.44
RBD4	99.1
SSM/RBD5	74.83
RBD2-RBD3	50.69
RBD3-RBD4	22.22
RBD4-TBD	9.57
TBD-SSM/RBD5	42.92

Table 3.3: Protein concentration used for NMR experiments.

3.2.1.8 Cryo-EM

Cryo-EM data for hStau1^{55}_FL, purified at 1.3 mg/ml in buffer A (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 200 mM L-Arg HCl) as per section 3.2.1.2, were collected by Dr G. Cannone (LMB, Cambridge) and analysed by Dr Laura Spagnolo. The low-resolution EM map for Staufen full-length was used to interpret SAXS scattering curves.

3.2.2 ARF1 RNA

3.2.2.1 ARF1 RNA preparation

ARF1 DNA was amplified by PCR from vector pCR2.1-TOPO_Arf1_SBS, gift of Dr Michael L. Gleghorn and Prof Lynne E. Maquat, using forward primer (5'-GTAATACGACTCACTATAGGGAGAGGCAAACGTGCGGCTCGTGG-3') and reverse primer (5'-GCTGAGTAATATTGCATAGG-3') as described in Table 3.4. The PCR product was purified from 0.8% agarose gel run in TBE using the MinElute Gel Extraction Kit (QIAGEN). The DNA was dried using a speed vacuum system and then resuspended in 30 µl of DEPC water. *In vitro* transcription was carried out for 6 hours at 37°C, supplementing the reaction mix with pyrophosphatase inorganic (NEB) and RNasin® Ribonuclease Inhibitors (Promega). ARF1 RNA was purified by phenol-chloroform extraction (3 washes with phenol:chloroform:isoamyl alcohol 25:24:1 followed by 3 washes with chloroform) and precipitated o/n at -20°C by addition of 2.5 volumes of EtOH and 0.1 volumes of 3 M potassium acetate. The following day, the sample was spun down at maximum speed in a pre-cooled bench-top centrifuge for 30 minutes at 4°C. The pellet was washed twice with 1 ml of 70% EtOH and then dried using a speed vacuum system. The RNA pellet was then resuspended in RNA storage buffer (40 mM HEPES, 100 mM KCl), aliquoted in PCR tubes, flash-frozen and stored at -80°C. Prior to use, RNA aliquots were thawed and boiled at 95°C for 10 minutes and then transferred into boiling water, allowing a slow and constant decrease in temperature. When the temperature was below 60°C, 1 mM $MgCl_2$ was added in order to favour RNA folding.

PCR step	Temperature (°C)	Time (sec)	
Initial denaturation	98	60	
Denaturation	98	10	41 cycles
Primers annealing	58	30	
Extension	72	15	
Final extension	72	120	
Hold	10	hold	

Table 3.4: PCR program used to amplify ARF1 DNA from vector pCR2.1-TOPO_Arf1_SBS.

3.2.2.2 Microscale thermophoresis (MST) of ARF1 RNA and Staufen proteins complexes

hStau1^{55}_FL and hStau1^{55}_ΔRBD2 were fluorescently labeled *via* the His-tag using NT-647. NT-647 labeled protein concentration was kept constant (10 nM), while the concentration of the non-labeled Arf1 RNA was varied between 33 µM and 1 nM. The samples were incubated for 30 minutes at room temperature and then loaded into MST NT.115 premium glass capillaries. MST analysis was performed using the Monolith NT.115Pico during a demonstration from NanoTemper.

3.2.2.3 Purification of Staufen proteins and Arf1 RNA complex

100 µl of hStau1^{55}_FL or hStau1^{55}_ΔRBD2 proteins (10 µM) were incubated with 100 µl of Arf1 RNA (87 µM) for 30 minutes at room temperature before being applied on top of a 24 ml glycerol gradient (5-70% glycerol in buffer 25 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 200 mM L-Arg HCl). hStau1^{55}_FL (10 µM),

hStau1^{55}_ΔRBD2 (10 µM), Arf1 RNA (87 µM) and gel filtration standards were individually loaded as controls. The gradient was run for 45 hours at 4°C at 28000 rpm on a Sorvall ultracentrifuge with swing-out rotor Surespin (Sorvall). The gradients were manually fractionated in fractions of 500 µl and separation was checked with SDS-PAGE analysis of 12% gels stained with SimplyBlue SafeStain (Invitrogen) and 0.8% agarose gel run in TBE.

3.2.2.4 Negative staining of hStau1^{55}_ΔRBD2/Arf1 complex

Negative staining EM and single particle analysis (SPA) were coupled to study hStau1^{55}_ΔRBD2/Arf1 complex. 5 µl of the fraction from the glycerol gradient purification containing both hStau1^{55}_ΔRBD2 protein and Arf1 RNA were deposited on an EM copper grid 400 mesh (Agar Scientific) covered with a ~10 nm layer of carbon and incubated for 2 minutes. The grid was washed and blotted 3 times with 10 µl of ultra-pure water and 3 times with 10 µl of 2% uranyl acetate. The grid was then incubated for 5 minutes with 5 µl of 2% uranyl acetate, blotted and then dried and stored at room temperature. The data collection was carried out at nominal magnification of 50,000x, at a final sampling of 1.9 Å/pixel at the specimen level, using a Jeol2200 200kV FEG microscope, equipped with a Gatan UltraScan 4000 (4096x4096) CCD camera. Data were processed using EMAN2 and RELION softwares. From the micrographs dataset, ~290,000 single particles (SP) were auto-picked with EMAN2 using 250 Å masks; the single particles images were extracted into boxes 128x128 pixels and band-pass filtered. The dataset was subjected to sequential rounds of alignments and classification in RELION, using a 200 Å mask, in order to polish the dataset and improve the resulting class-average images. ~43,000 particles were then sorted with RELION into three 3D classed using as reference an *ab initio* initial model generated with EMAN2. Figures were prepared with UCSF Chimera.

3.2.3 Upf1 protein

3.2.3.1 Cloning of hUpf1

Oligos containing the sequence for STREP-tag (Table 3.5) were annealed to form dsDNA: 50 µl of each oligos (100 µM) were added into a PCR tube and mixed before being boiled at 95°C for 10 minutes, annealing was achieved o/n by incubation of the PCR tubes into boiling water, allowing a slow and constant decrease in temperature. dsDNA containing Strep-tag and pTWO-E vector were digested with XbaI and HindIIIHF (NEB) for 2 hours at 37°C. Digested vector was purified from 1% agarose gel run in TBE using the MinElute Gel Extraction Kit (QIAGEN), whereas digested Strep-tag insert was purified using the QIAquick PCR Purification Kit (QIAGEN). Ligations between the vector and the insert (in ratio 1:3) were performed using the Quick Ligation Kit (NEB) for 5 minutes at RT. *E. coli* XL1-Blue cells were transformed by heat-shock method with 2.5 µl of the ligation reaction and plated in LB agar plates containing 50 µg/ml ampicillin. After overnight, 15 ml of LB, supplemented with 50 µg/ml ampicillin, were inoculated with a single colony and the cultures were grown at 37°C o/n. Subsequently, plasmid DNA was purified using the Wizard Plus SV Minipreps DNA Purification Systems and sequenced. pTWO-E_Strep-tag and pGEX-hUpf1 vectors were digested with NotIHF and HindIIIHF (NEB) for 3 hours at 37°C. Digested hUPF1 insert and pTWO-E_Strep-tag vector were purified from 1% agarose gel run in TBE using the MinElute Gel Extraction Kit (QIAGEN). Ligations between the vector and the insert, transformation and plasmid DNA extraction were performed as described above and sequenced.

OLIGO name	Sequence
STREPtag_SENSE	5'-TTTTTTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATAAGATCT ATGTGGAGCCACCCGCAGTTCGAGAAAAAGCTTTTTTTTT-3'
STREPtag_ANTISENSE	5'-AAAAAAAAGCTTTTTCTCGAACTGCGGGTGGCTCCACATAGATCTTATGTATAT CTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAAAAAAA-3'

Table 3.5: Oligos used to introduce Strep-tag sequence in pTWO-E vector using XbaI and HindIIIHF restriction enzymes.

3.2.3.2 hUpf1 overexpression and purification

Recombinant hUpf1 protein was overexpressed in *E. coli* Rosetta™(DE3) pLysS cells transformed by heat-shock with pTWO-E_Strep-tag vectors containing hUpf1, which were obtained as described in 3.2.2.1. Starting culture was grown in LB medium containing 50 µg/ml ampicillin and 34 µg/ml chloramphenicol at 37°C o/n. 4 ml of overnight culture was inoculated in 4 L of TB medium supplemented with antibiotics and cells were grown to an OD_{600}=0.6. Protein overexpression was induced by adding 0.25 mM IPTG and culturing the cells at 23°C for 4 hours. Cells were harvested by centrifugation at 5000 rpm on a Beckman Avanti™ J-20 XP centrifuge equipped with JLA 8.1000 rotor for 20 minutes and washed once with PBS; cell pellets were aliquoted and stored at -80°C. Four frozen aliquots, from 1 L of culture each, were thawed and lysed by sonication in 30 ml of lysis buffer (25 mM HEPES pH 7.4, 1 M NaCl, 1mM $MgCl_2$, 1µM $ZnCl_2$, 20 mM imidazole, 1% Triton X-100, 600 µl Complete EDTA-free protease inhibitors 50X). Soluble protein extracts were separated from cell pellets by centrifugation at 15000 rpm for 30 min. 6-His tagged hUpf1 protein was purified by nickel chromatography using HisTrap FF columns (GE Healthcare) equilibrated in washing buffer B (25 mM HEPES pH 7.4, 1 M NaCl, 1 mM $MgCl_2$, 1 µM $ZnCl_2$, 20 mM imidazole). Elution was performed with a 20 mM-1 M imidazole gradient. Fractions containing the protein of interest were pooled and incubated o/n at 4°C with 2 ml of streptactin resin, previously equilibrated with buffer B. The following day, the resin was transferred to a gravity column and washed with 5CV of buffer B. Elution was performed in buffer B supplemented with 2.5 mM desthiobiotin. The elution from the streptactin resin was concentrated to 500 µl [2 mg/ml] and loaded into a Superdex 200 Increase 10/300 GL SEC column, previously equilibrated in buffer C (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$). Overexpression and purification efficiency were monitored by SDS-PAGE analysis on 8% gels stained with SimplyBlue SafeStain (Invitrogen). Fractions were also analyzed by Western Blot: the membrane was incubated o/n at 4°C with anti-His antibody conjugated with alkaline phosphatase (1:4000) and developed using the SIGMA-FAST™ BCIP/NBT reagent.

3.2.3.3 Interaction between Staufen proteins, Arf1 RNA and hUpf1 protein

70 µM of purified hStau1^{55}_FL or hStau1^{55}_ΔRBD2 proteins were firstly incubated with 70 µM of Arf1 RNA for 30 minutes at room temperature. 70 µM of purified hUpf1 were then added to the reaction mix and incubated at room temperature for 1 hour, in a final volume of 21 µl in buffer C (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$). Control reactions were set up with single components and protein-protein and protein-RNA pairs. 7.5 µl of each reaction was loaded on a 4-16% Bis-Tris NativePAGE (Invitrogen) with 0.5 µl of NativePage 5% G-250 sample additive (Invitrogen). Dark blue BN-PAGE was run for 1 hour at 150 V. Gel was de-stained o/n in 30% ethanol and 10% acetic acid.

3.3 Results and Discussion

3.3.1 Staufen proteins

3.3.1.1 Solution studies of Staufen domains

Individual and tandem hStau1[55] domains were purified to homogeneity by immobilized nickel chromatography followed by size exclusion chromatography using a Superdex 75 Increase 10/300 GL column. The SEC traces and SDS-PAGE analysis of individual domains are shown in Figure 3.2A-B, while the ones for tandem domains are shown in Figure 3.2C-D. For each construct, SEC peaks are consistent with the molecular weight of a monomer and correspond to ~10-12 KDa for individual domains and to ~20-24 KDa for tandem domains.

Interestingly, a higher molecular weight species can be observed in the SEC trace of RBD3 and RBD3-RBD4 and, in smaller amount, in the one of RBD4. SDS-PAGE analysis and $A_{260/280}$ measurements confirmed these species to be nucleic acid contaminants, previously co-purified during the affinity chromatography step. Their removal was successfully achieved by buffer exchange. Fractions from individual and tandem domains SEC peaks were concentrated to ~30 μM using 3 K MWCO spin-concentrators and analytical ultracentrifugation (AUC) experiments were performed in order to address whether or not these constructs could form dimers in solution (Figure 3.2E-F).

Our data for the SSM/RBD5 construct are consistent with that shown previously by X-ray crystallography (Gleghorn et al., 2013) and show that SSM/RBD5 is present in solution both as a monomer and as a dimer, likely to be a concentration-dependent equilibrium. However, TBD seems to inhibit the dimerization of SSM/RBD5, as supported by the presence of a unique peak in the AUC data for the tandem domain TBD-SSM/RBD5. Moreover, the construct RBD3-RBD4 is mainly present in solution as a monomer, as previously shown (Lazzaretti et al., 2018), but has a low tendency to form dimers even in the absence of a RNA substrate.

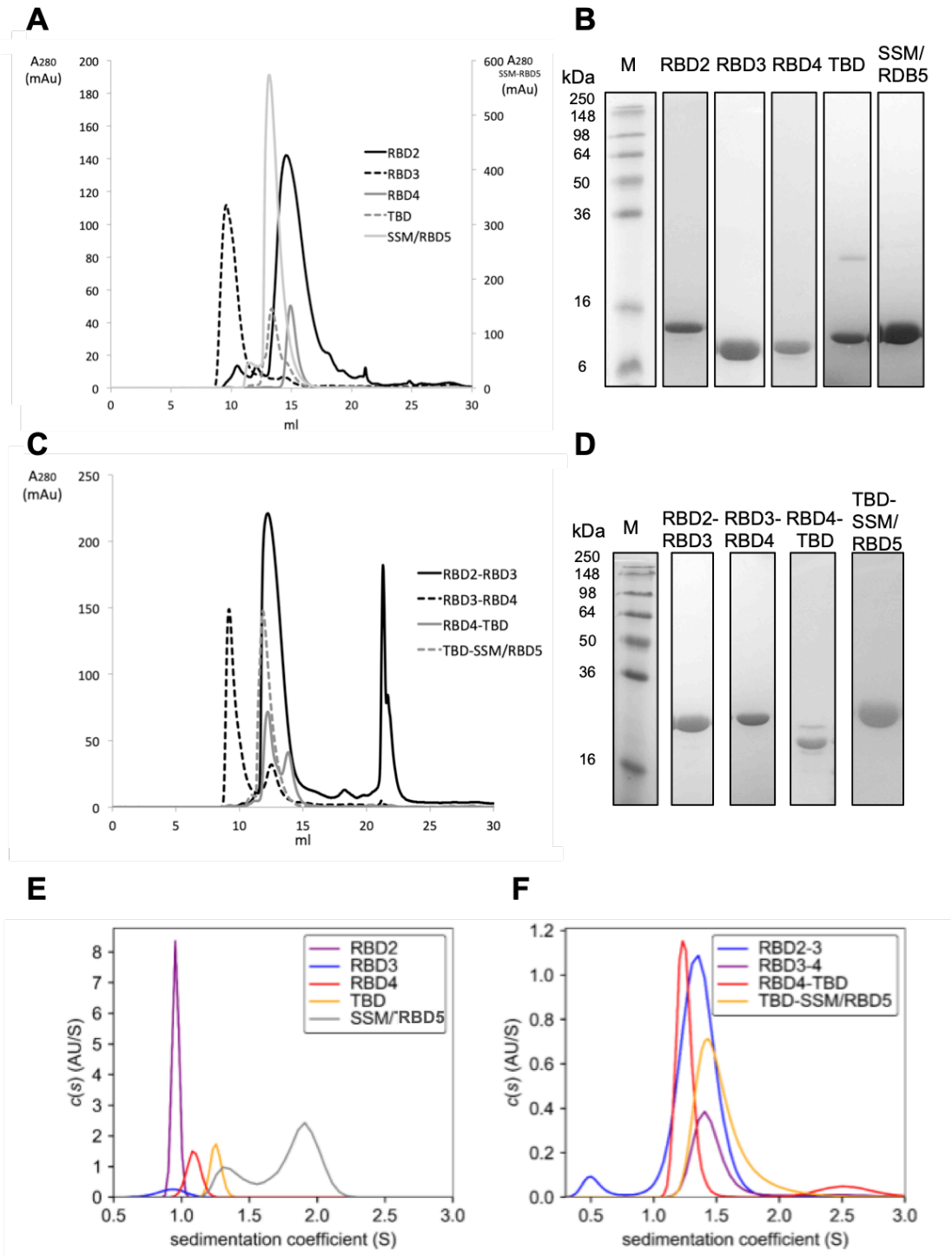


Figure 3.2: Purification and biophysical characterization of Staufen individual and tandem domains. A: SEC trace of Staufen individual domains purified in buffer A (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM MgCl2, 200 mM L-Arg HCl) using a Superdex 75 Increase 10/300 GL SEC column at 0.7 ml/min. B: SimplyBlue SafeStain staining of 18% SDS-PAGE gel of SEC purified individual domains. C: SEC trace of Staufen tandem domains purified in buffer A (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM MgCl2, 200 mM L-Arg HCl) using a Superdex 75 Increase 10/300 GL SEC column at 0.7 ml/min. D: SimplyBlue SafeStain staining of 18% SDS-PAGE gel of SEC purified tandem domains. E: Analytical ultracentrifugation (AUC) of individual domains (~30 µM). F: Analytical ultracentrifugation (AUC) of tandem domains (~30 µM).

Individual and tandem domains were studied by SAXS and 2D NMR (Figure 3.3 and Figure 3.4, respectively). $^{15}N,^{1}H$-TROSY spectra were recorded for individual and tandem domains, with the exception of TBD, for which the concentration was too low.

The presence of peaks only in the central area of the spectra of RBD2 indicates that this domain is unfolded in isolation. RBD3 and RBD4 are folded domains in isolation and peaks corresponding to amino acids in regions with a well-defined secondary structure and the ones corresponding to amino acids in disordered regions (loops) can be seen in their NMR spectra. SSM/RBD5 construct appears to be structured, however, some buffer contribution and possibly monomer-dimer equilibrium interfere with further investigations.

Additionally, the similar lineshapes and lack of significant chemical shift perturbations between the spectra recorded for individual and tandem domains (Figure 3.3 and 3.4, respectively) show that interactions between adjacent domains are very limited and that each of the domains constituting the tandem constructs tumbles independently in solution, thanks to the intervening linkers' flexibility. This is also confirmed by SAXS analysis of individual and tandem domains, with representative models shown appended to NMR spectra. SAXS profiles, Kratky analysis and the ensemble of models obtained for each construct are shown in Figures 3.5-3.13.

Different modelling approaches were adopted based on the flexibility level of each construct. RBD2 was modeled using EOM in "compact chain" mode, out of 10,000 models initially generated based on its sequence, seven were selected as final output that satisfy the experimental data and the dimensional parameters of the models are R_g*ensemble*=37.3 Å**,** D_{max}*ensemble*=112.64 Å, (X^2=0.98). RBD3 and RBD4 were modeled using DAMMIN as they appeared to be more compact. RBD3 dimensions are R_g=18.74 Å and D_{max}=58.08 Å (X^2=0.835), while RBD4 has R_g=17.15 Å and D_{max}=51.37 Å (X^2=0.5314). The same approach used for the unstructured RBD2 in isolation was used for TBD and EOM generated eight models (R_g*ensemble*=31.74 Å**,** D_{max}*ensemble*=97.24 Å (X^2=1.29). The other domains were modeled using either EOM or BUNCH and feeding the software homology models obtained for RBD3, RBD4 and SSM/RBD5 as constraints. Four models were generated for SSM/RBD5 (R_g*ensemble*=25.55 Å**,** D_{max}*ensemble*=83.44 Å (X^2=1.405), for RBD2-RBD3 (R_g*ensemble*=37.86 Å**,**

D_{max}*ensemble*=120.74 Å (X^2=1.16) and for RBD3-RBD4 (R_g*ensemble*=36.15 Å, D_{max}*ensemble*=114.96 Å (X^2=1.342). Eight models were generated for TBD-RBD4 (R_g*ensemble*=36.46 Å, D_{max}*ensemble*=121.16 Å (X^2=1.253). The model obtained with BUNCH for TBD-SSM/RBD5 shows a R_g=41.24 Å and a D_{max}=129.65 Å (X^2=1.28).

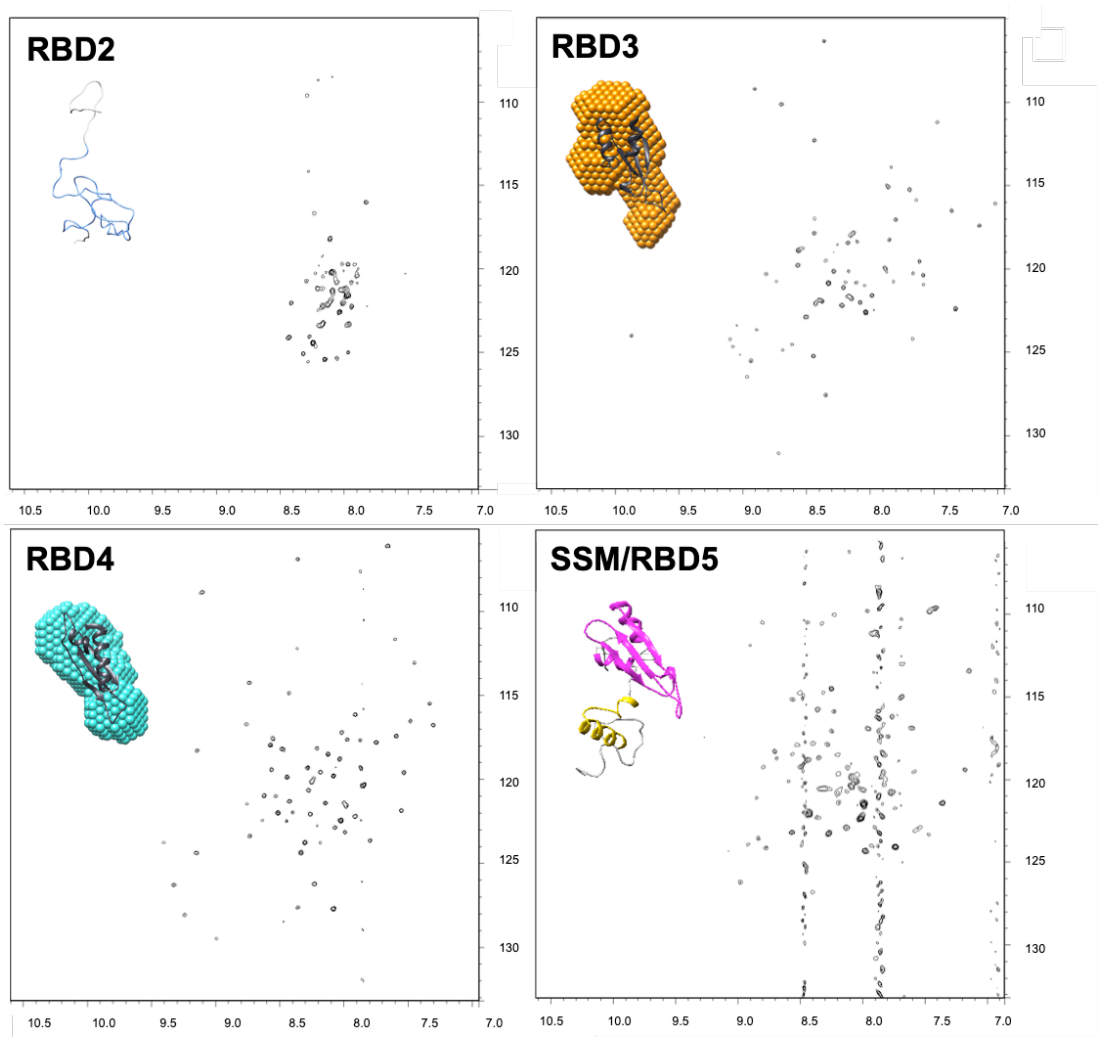


Figure 3.3: NMR of Staufen individual domains and representative SAXS models. For NMR experiments, 30 µl of D_2O were added to 570 µl of each protein in 20 mM potassium phosphate pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 200 mM L-Arg.HCl, 2 mM TCEP and ^{15}N,^{1}H-TROSY-HSQC spectra were acquired at 298 K using a Bruker AVANCE IIIHD 600 MHz spectrometer equipped with a 5 mm TCI cryoprobe. Data were processed using the Bruker TopSpin software and figures were generated using CCPN analysis 2.4. SAXS data collected at B21, Diamond Light Source (Harwell, UK). 55 µl of each protein sample (~10 mg/ml) was loaded onto a Superdex 75 Increase 10/300 GL column (GE Healthcare), controlled by an Agilent HPLC system, coupled to an in-vacuum SAXS flow cell and HPLC-SAXS traces were processed using ScÅtter. RBD3 and RBD4 were modelled with DAMMIN, RBD2 and SSM/RBD5 with EOM.

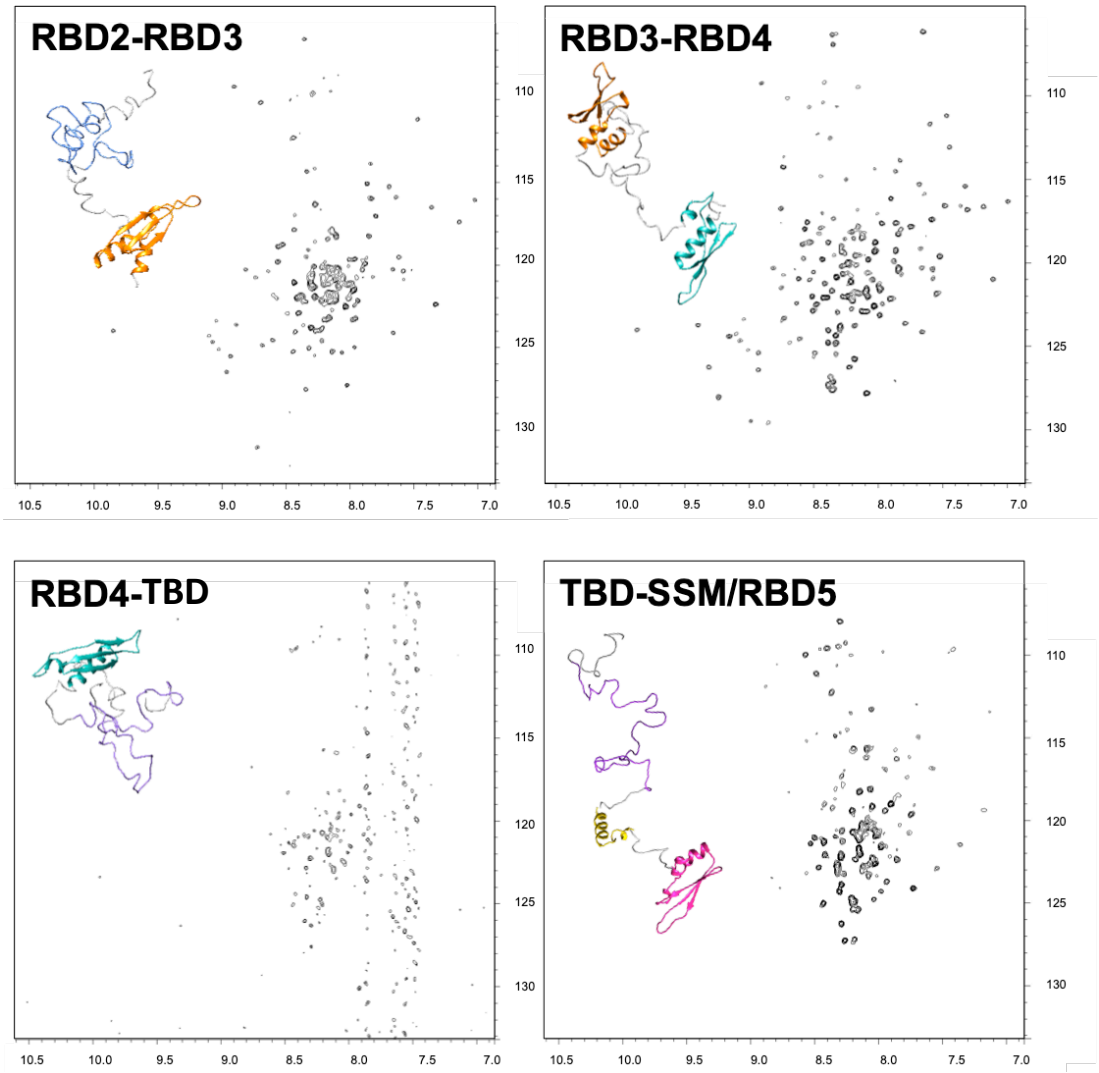


Figure 3.4: NMR of Staufen tandem domains and representative SAXS models. For NMR experiments, 30 µl of D_2O were added to 570 µl of each protein in 20 mM potassium phosphate pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 200 mM L-Arg.HCl, 2 mM TCEP and ^{15}N,^{1}H-TROSY-HSQC spectra were acquired at 298 K using a Bruker AVANCE IIIHD 600 MHz spectrometer equipped with a 5 mm TCI cryoprobe. Data were processed using the Bruker TopSpin software and figures were generated using CCPN analysis 2.4. SAXS data collected at B21, Diamond Light Source (Harwell, UK). 55 µl of each protein sample (~10 mg/ml) was loaded onto a Superdex 75 Increase 10/300 GL column (GE Healthcare), controlled by an Agilent HPLC system, coupled to an in-vacuum SAXS flow cell and HPLC-SAXS traces were processed using ScÅtter. TBD-SSM/RBD5 was modelled with BUNCH and all the other tandem domains with EOM.

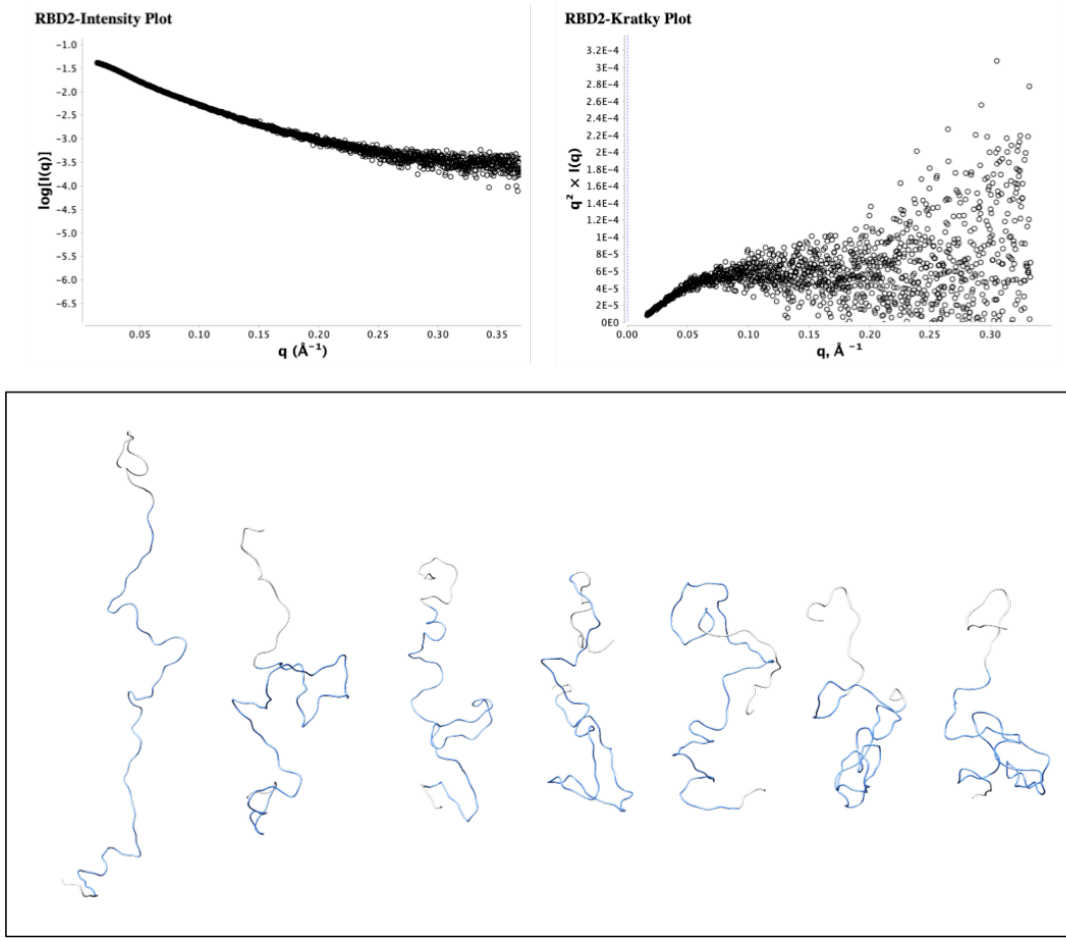


Figure 3.5: Small angle X-ray scattering (SAXS) of RBD2. SAXS curve, Kratky analysis and EOM models generated for RBD2. R_g *ensemble*= 37.3 Å, D_{max} *ensemble*= 112.64 Å, X^2=0.98, R_{sigma}= ~85.78% (~86.70%).

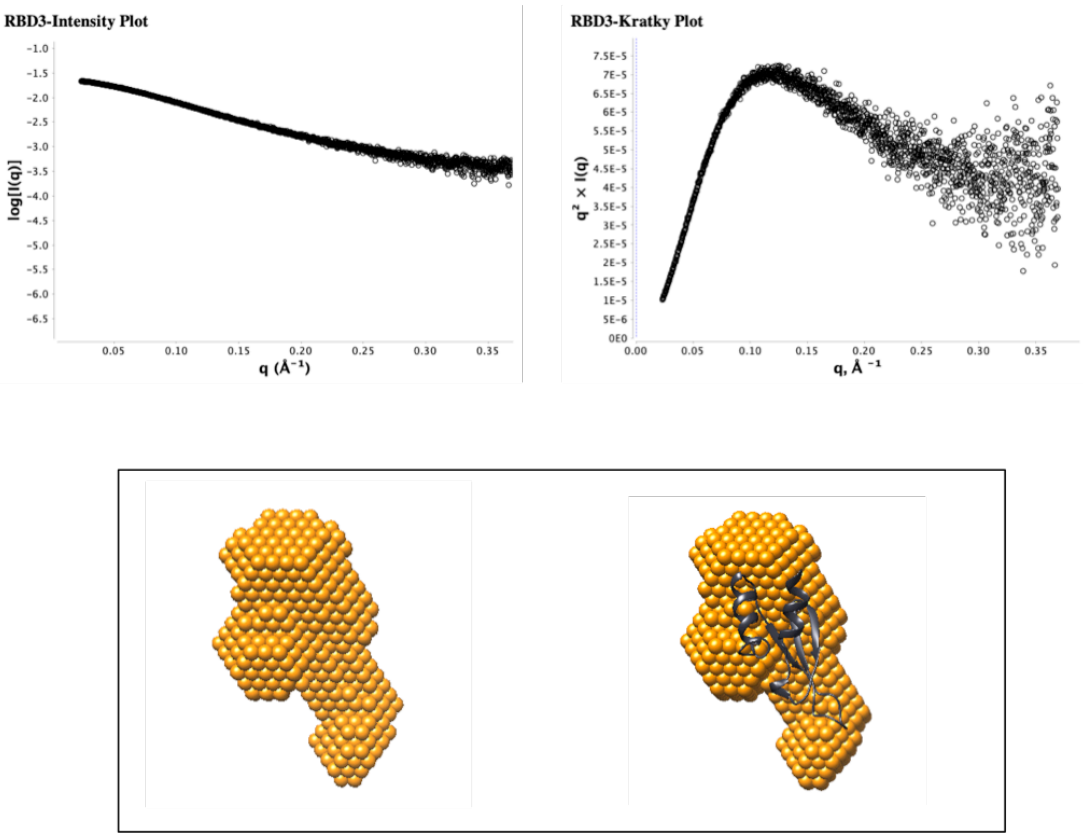


Figure 3.6: Small angle X-ray scattering (SAXS) of RBD3. SAXS curve, Kratky analysis and DAMMIN models generated for RBD3. R_g= 18.74 Å, D_{max}= 58.08 Å, X^2= 0.835.

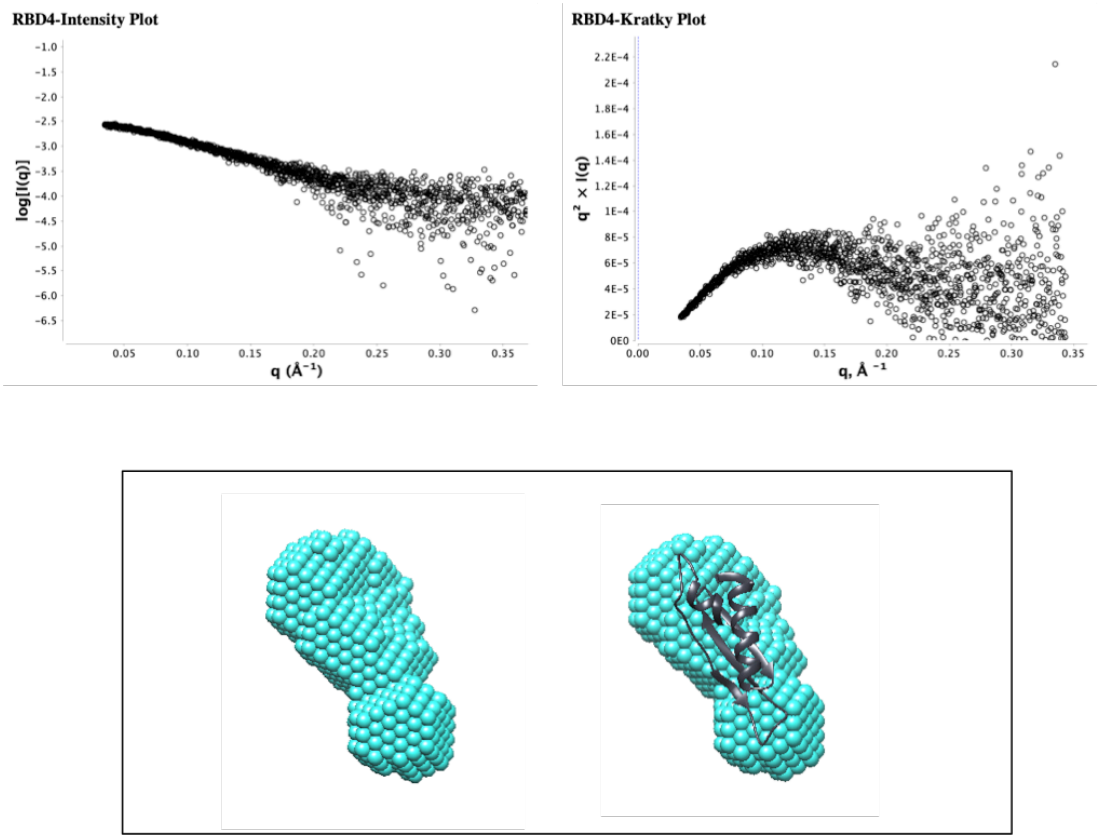


Figure 3.7: Small angle X-ray scattering (SAXS) of RBD4. SAXS curve, Kratky analysis and DAMMIN models generated for RBD4. R_g= 17.15 Å, D_{max}= 51.37 Å, X^2= 0.5314.

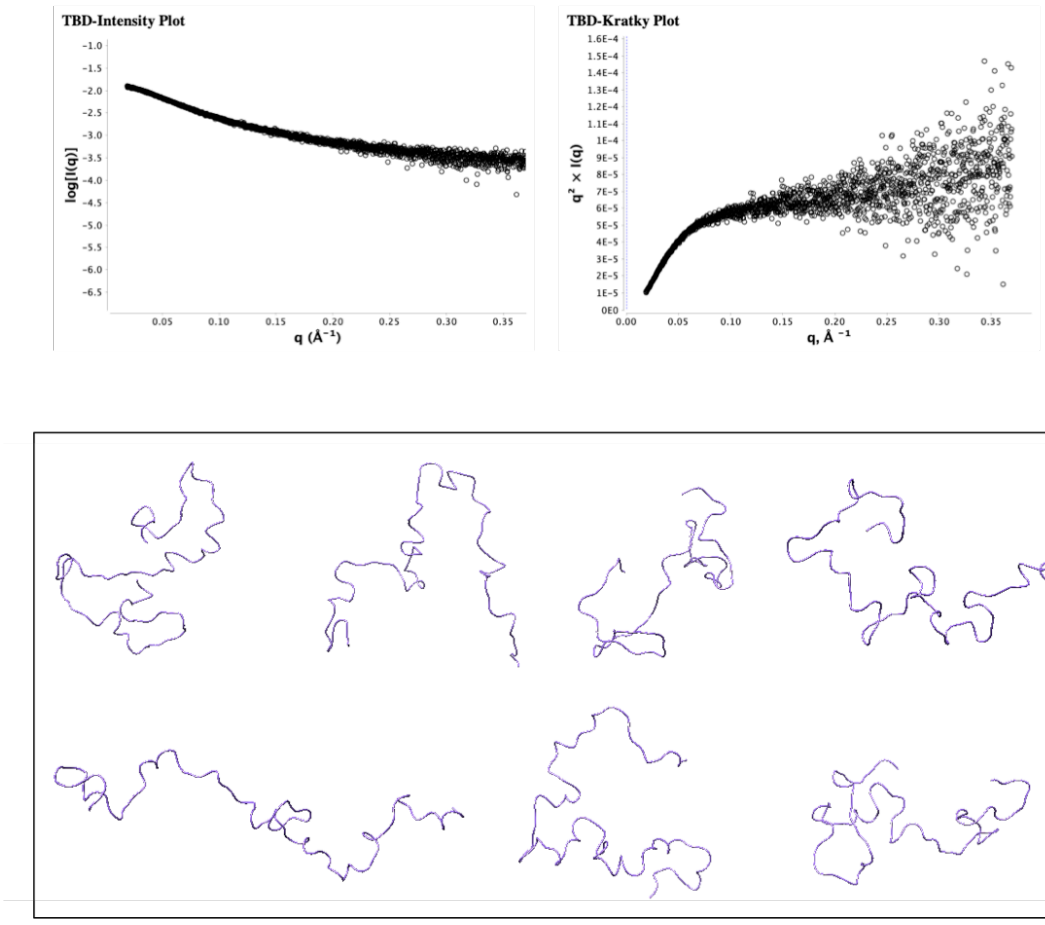


Figure 3.8: Small angle X-ray scattering (SAXS) of TBD. SAXS curve, Kratky analysis and EOM models generated for TBD. R_g *ensemble*= 31.74 Å, D_{max} *ensemble*= 97.24 Å, X^2= 1.29, R_{sigma}= ~88.51% (~86.24%).

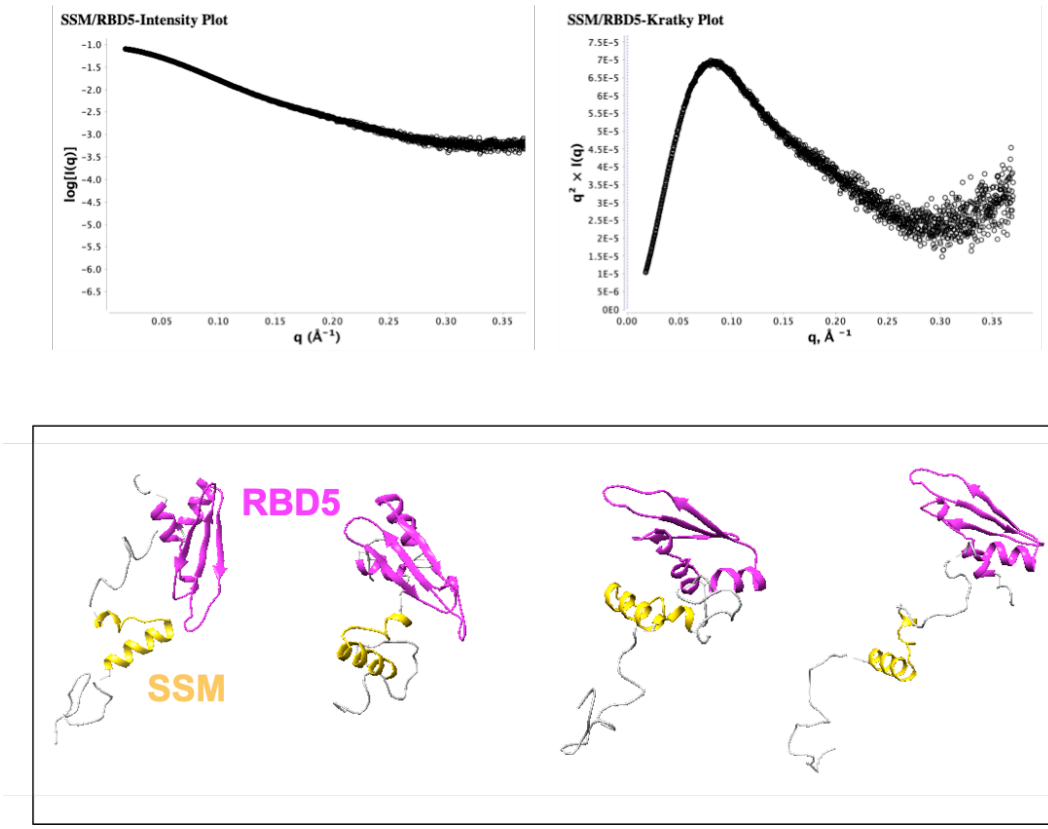


Figure 3.9: Small angle X-ray scattering (SAXS) of SSM/RBD5. SAXS curve, Kratky analysis and EOM models generated for SSM/RBD5. R_g *ensemble*= 25.55 Å, D_{max} *ensemble*= 83.44 Å, X^2= 1.405, R_{sigma}= ~81.16% (~85.66%).

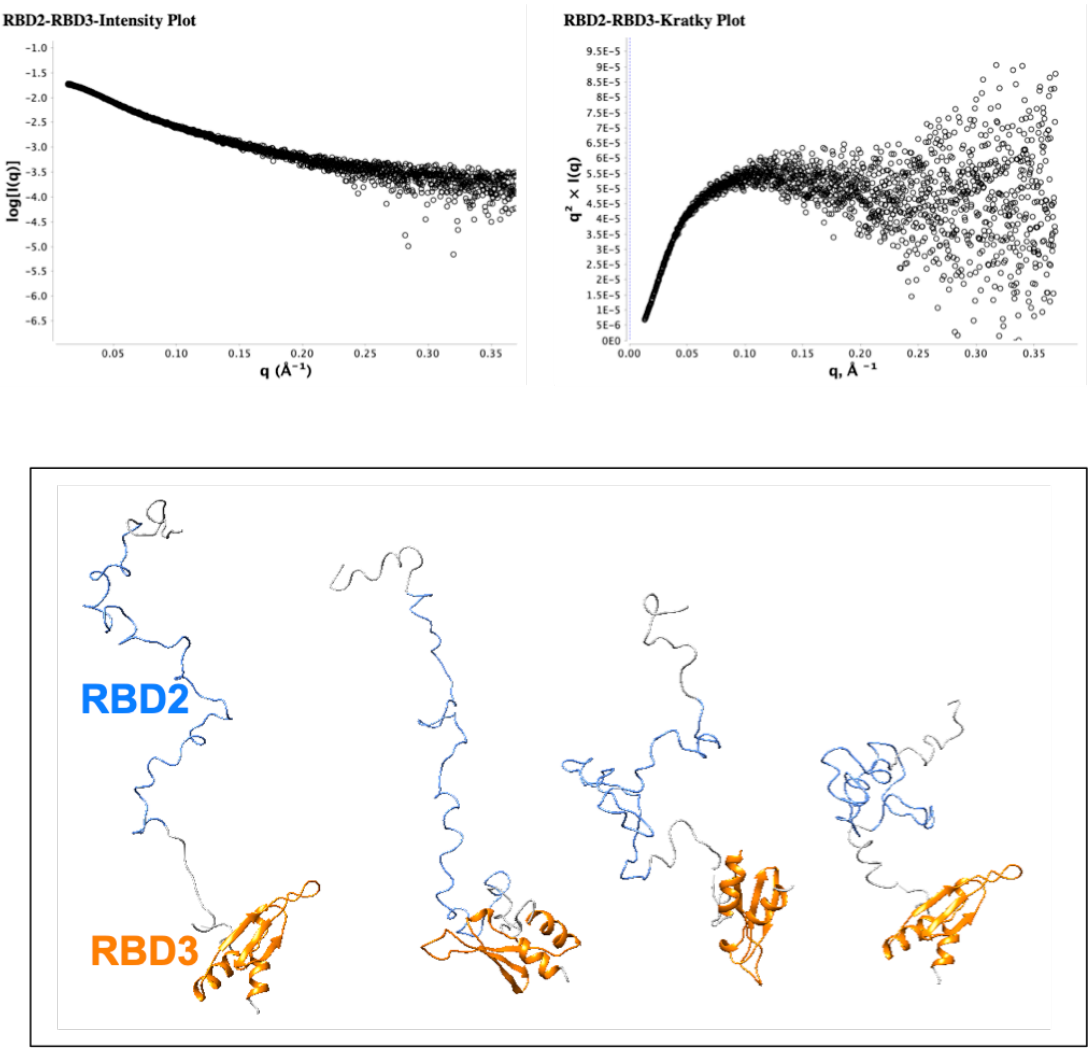
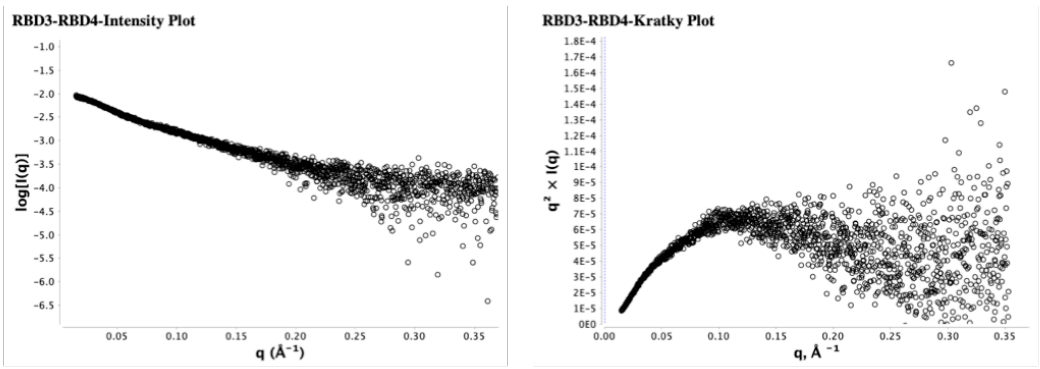


Figure 3.10: Small angle X-ray scattering (SAXS) of RBD2-RBD3. SAXS curve, Kratky analysis and EOM models generated for RBD2-RBD3. R_g *ensemble*= 37.86 Å, D_{max} *ensemble*= 120.74 Å, X^2= 1.16, R_{sigma}= ~85% (~85.85%).



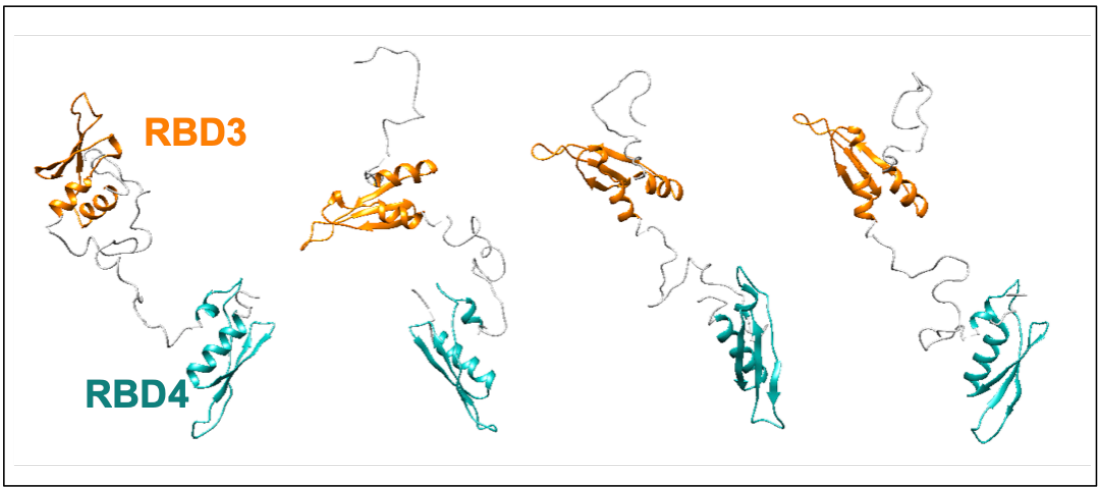
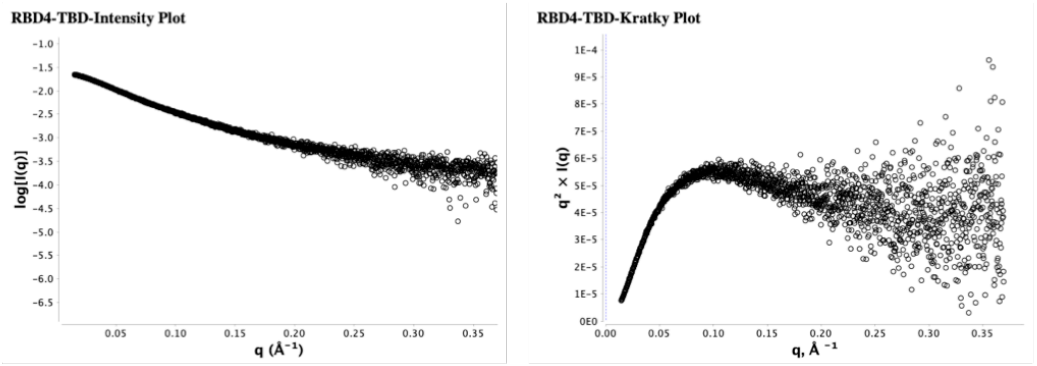


Figure 3.11: Small angle X-ray scattering (SAXS) of RBD3-RBD4. SAXS curve, Kratky analysis and EOM models generated for RBD3-RBD4. R_g *ensemble*= 36.15 Å, D_{max} *ensemble*= 114.96 Å, X^2= 1.342, R_{sigma}= ~80.28% (~85.03%).



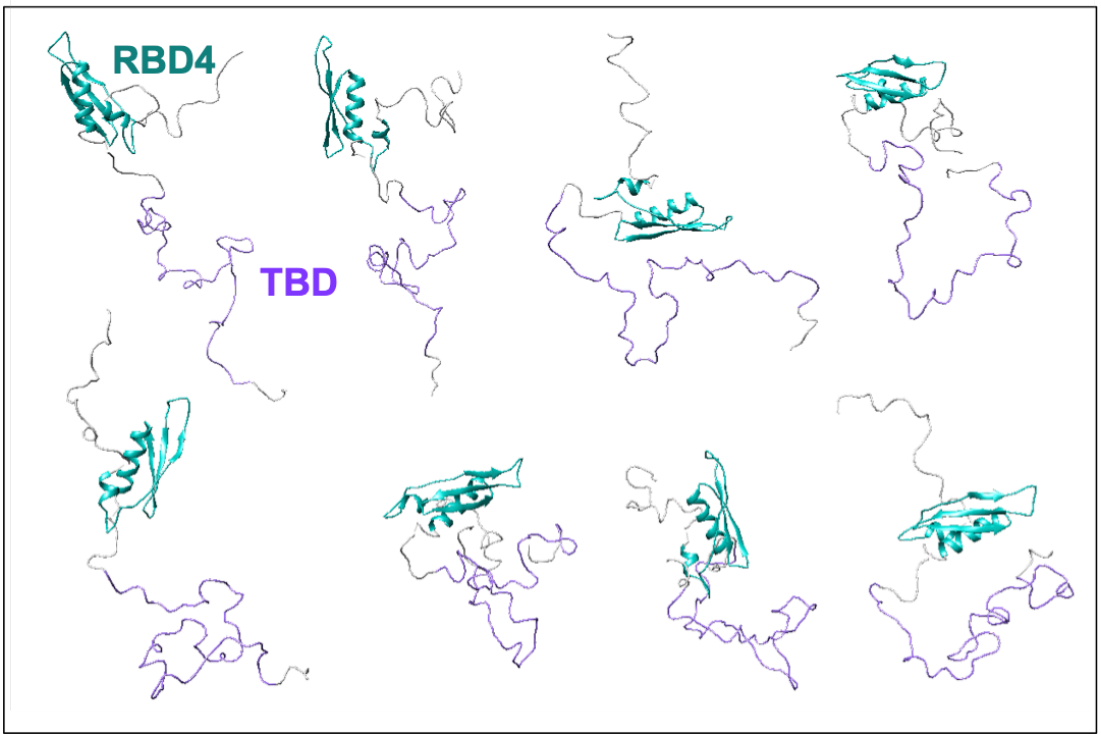


Figure 3.12: Small angle X-ray scattering (SAXS) of RBD4-TBD. SAXS curve, Kratky analysis and EOM models generated for RBD4-TBD. R_g *ensemble*= 36.46 Å, D_{max} *ensemble*= 121.16 Å, X^2= 1.253, R_{sigma}= ~88.79% (~86.24%).

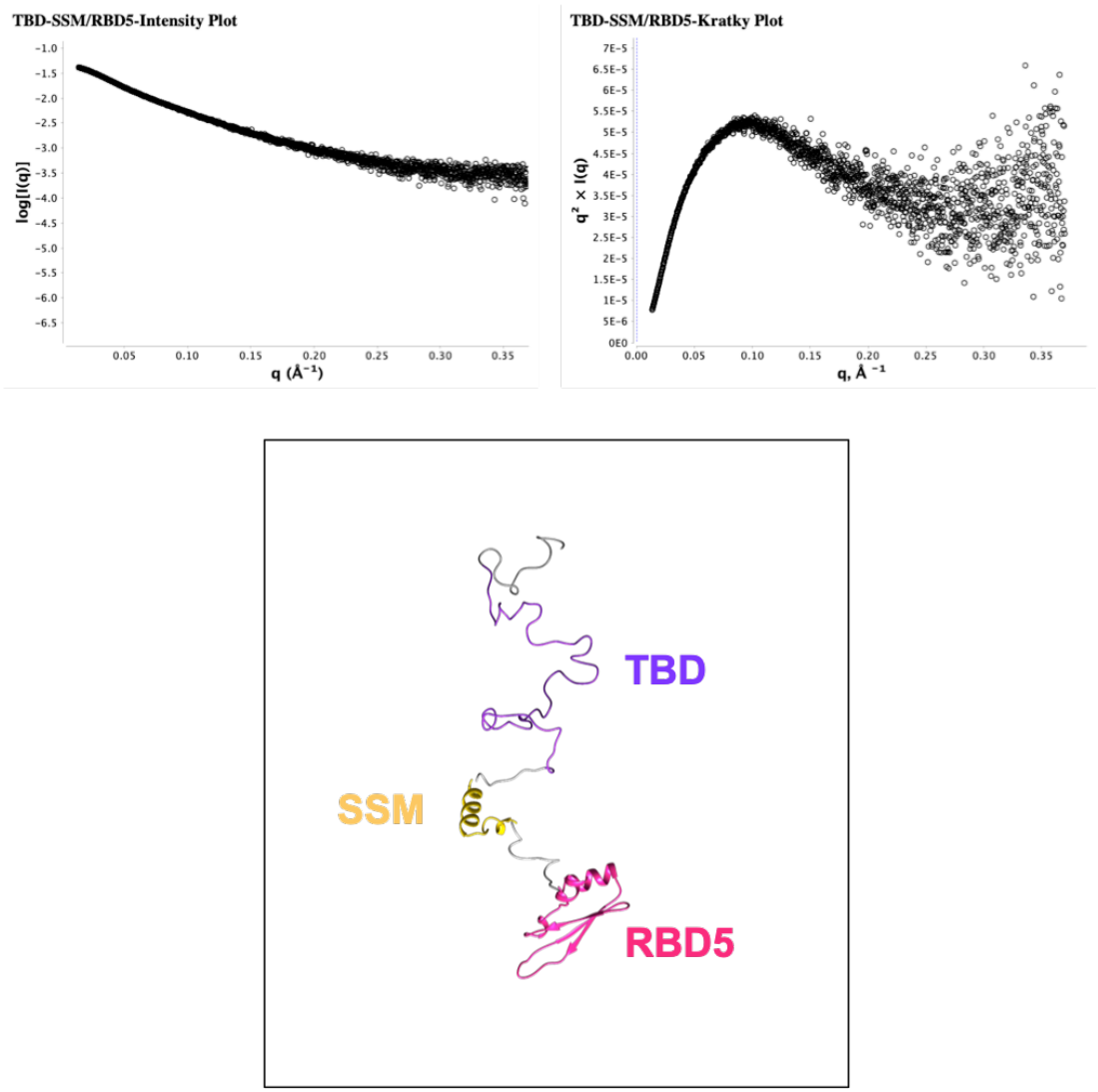


Figure 3.13: Small angle X-ray scattering (SAXS) of TBD-SSM/RBD5. SAXS curve, Kratky analysis and BUNCH models generated for TBD-SSM/RBD5. R_g *ensemble*= 41.24 Å, D_{max} *ensemble*= 129.65 Å, X^2= 1.28, R_{sigma}= ~89.35% (~85.53%).

3.3.1.2 RBD2 influences both solubility and oligomeric state of Staufen1

Staufen proteins were purified to homogeneity by immobilized nickel chromatography followed by size exclusion chromatography.

Additive screening to discover conditions that would allow the protein to achieve high concentration and good homogeneity for subsequent structural studies was performed using 10 K MWCO spin-concentrators (Rambo, 2017). The addition of L-Arg HCl to the buffer proved necessary for maintaining the solubility of the full-length protein to enable further experiments. Interestingly, the solubility of hStau1^{55}_ΔRBD2 is not affected by the presence (or absence) of L-Arg HCl in the buffer. However, this additive was used for all constructs for consistency with the purification requirements of hStau1^{55}_FL.

Size exclusion chromatography (SEC) traces show that the hydrodynamic volumes of hStau1^{55}_FL and of hStau1^{55}_ΔRBD2 (~130-140 kDa) are higher than expected for globular proteins with corresponding molecular weights, suggesting that the two proteins might have an elongated shape or might form homo-multimers (Figure 3.14).

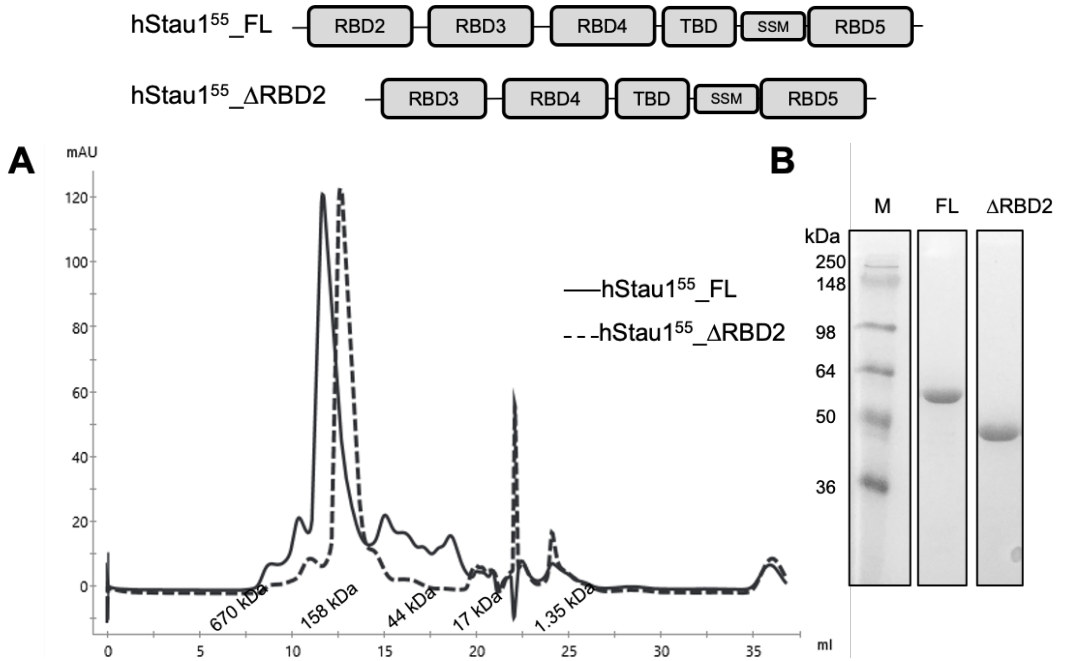


Figure 3.14: Domains organization and purification of hStau1^{55}_FL and hStau1^{55}_ΔRBD2. A: Size exclusion chromatography trace of hStau1^{55}_FL (continuous line) and of hStau1^{55}_ΔRBD2 (dashed line) in buffer A (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM MgCl2, 200 mM L-Arg HCl) using a Superdex 200 Increase 10/300 GL column at 0.5 ml/min flow-rate (BioRad GF markers appended to the chromatogram). B: SimplyBlue staining of 12% SDS-PAGE analysis of purified hStau1^{55}_FL and hStau1^{55}_ΔRBD2.

SEC-MALLS analysis of hStau1^{55}_FL (Figure 3.15A) shows the presence of multiple assemblies. In contrast to the recently published SEC profile for hStau1^{63}_FL (Lazzaretti et al., 2018), both SEC profiles for hStau1^{55}_FL and hStau1^{55}_ΔRBD2 described in this study present symmetrical peaks.

MALLS analysis of the eluting species highlights the different behaviour of hStau1^{55}_FL and hStau1^{55}_ΔRBD2. Measurements were performed at three different concentrations (20, 100 and 200 μM). The samples used for this analysis had not been subject to the final SEC purification, hence small amounts of species other than hStau1^{55}_FL and hStau1^{55}_ΔRBD2 were seen to be present, however, the predominant peak in the hStau1^{55}_ΔRBD2 sample had a molecular weight consistent with that of a monomer (Figure 3.15B). The main SEC peak for hStau1^{55}_FL appears with a molecular weight consistent with the one of a dimer, but the high poly-dispersity seen across the peak suggests that this is an equilibrium species been a monomer and higher order oligomers. This is consistent with the previous observation that RBD2 mediates hStau1^{55}_FL self-association (Martel et al., 2010, Lazzaretti et al., 2018), showing that its presence is fundamental for the formation of a stable oligomer in solution. The recent report describing hStau1^{63} did not contain SEC-MALLS analysis for the full-length protein (Lazzaretti et al., 2018).

To resolve the oligomeric assemblies of the species eluting in the main SEC peak, analytical ultracentrifugation (AUC) experiments were performed. AUC analysis of the peak fraction from hStau1^{55}_FL SEC (Figure 3.15C) confirms the co-existence of a number of species with molecular weights consistent with that of the monomer (major species in solution) and higher oligomers. Measurements were performed at three different concentrations (4, 25 and 100 μM). We chose to analyse a wide range of concentrations to address the role of concentration in the oligomeric state of the protein. The number of oligomers increases with increasing sample concentration, and the position of the peaks also shifts to a higher sedimentation co-efficient. Both of these phenomena indicate concentration-dependent self-association equilibrium for hStau1^{55}_FL.

On the other hand, both SEC-MALLS (Figure 3.15B) and AUC (Figure 3.15D) analysis of hStau1^{55}_ΔRBD2 show that the truncated protein is only present in solution as monomer.

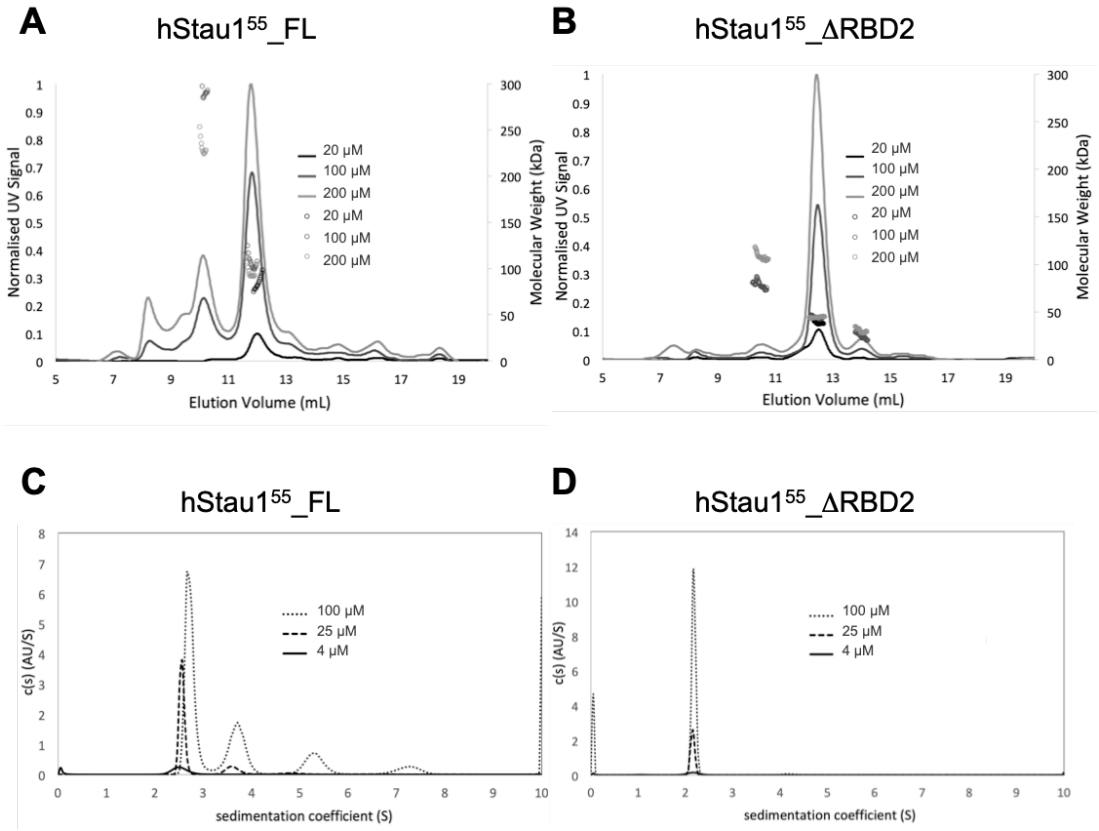


Figure 3.15: Biophysical characterization of Staufen proteins. A: SEC-MALLS of hStau1^{55}_FL at 20,100 and 200 µM. B: SEC-MALLS of hStau1^{55}_ΔRBD2 at 20,100 and 200 µM. C: Analytical ultracentrifugation (AUC) of hStau1^{55}_FL at 4, 25 and 100 µM. D: Analytical ultracentrifugation (AUC) of hStau1^{55}_ΔRBD2 at 4, 25 and 100 µM.

3.3.1.3 Staufen adopts distinct elongated structures in solution

The first indication that Staufen protein and its hStau1^{55}_ΔRBD2 mutant adopt elongated structures in solution is given by their average hydrodynamic volume, which is much higher than expected for the estimated molecular weight of the monomeric protein. This is in agreement with the R_g and D_{max} obtained from SAXS measurements for both the full-length (R_g*ensemble*= 48.11 Å, D_{max}*ensemble*= 155.2 Å) and truncated ΔRBD2 (R_g*ensemble*= 50.26 Å, D_{max}*ensemble*= 166.96 Å) proteins. The higher R_g and D_{max} for hStau1^{55}_ΔRBD2 can be explained by a higher degree of conformational heterogeneity in the FL protein. Importantly, SAXS shows that both these systems are characterized by a high degree of flexibility, as shown by their Kratky plot in Figure 3.16.

The three-dimensional models of hStau1^{55}_FL protein and of its truncation mutant hStau1^{55}_ΔRBD2 were obtained by combining homology modelling analysis and small angle X-ray scattering data.

The models obtained for hStau1^{55}_ΔRBD2 show that the protein adopts a range of conformations (Figure 3.16A), from highly extended to more compact, where RBD3 and RBD5 are in closer proximity. The relative positions of RBD4, TBD and SSM show only minor differences among the models obtained, due to the flexibility of the loops. On the other hand, the high level of flexibility of the loops between RBD3 and RBD4 and between SSM and RBD5 seems to be the main factor that contributes to the co-existence of a more distended and a more closed conformation of hStau1^{55}_ΔRBD2. These models show that all the individual domains do not coalesce to form a compact structure.

The models obtained for hStau1^{55}_FL, represented in Figure 3.16B, show more inter-domain flexibility, resulting in the presence of elongated, as well as more compact, conformations. The major differences between the co-existing conformations are due to the disordered loops between RBD2 and RBD3, between RBD3 and RBD4 and between RBD4 and TBD. Thanks to the malleability of these linkers, hStau1^{55}_FL seems to be able to transition in solution from more elongated to more closed conformations, but still not globular.

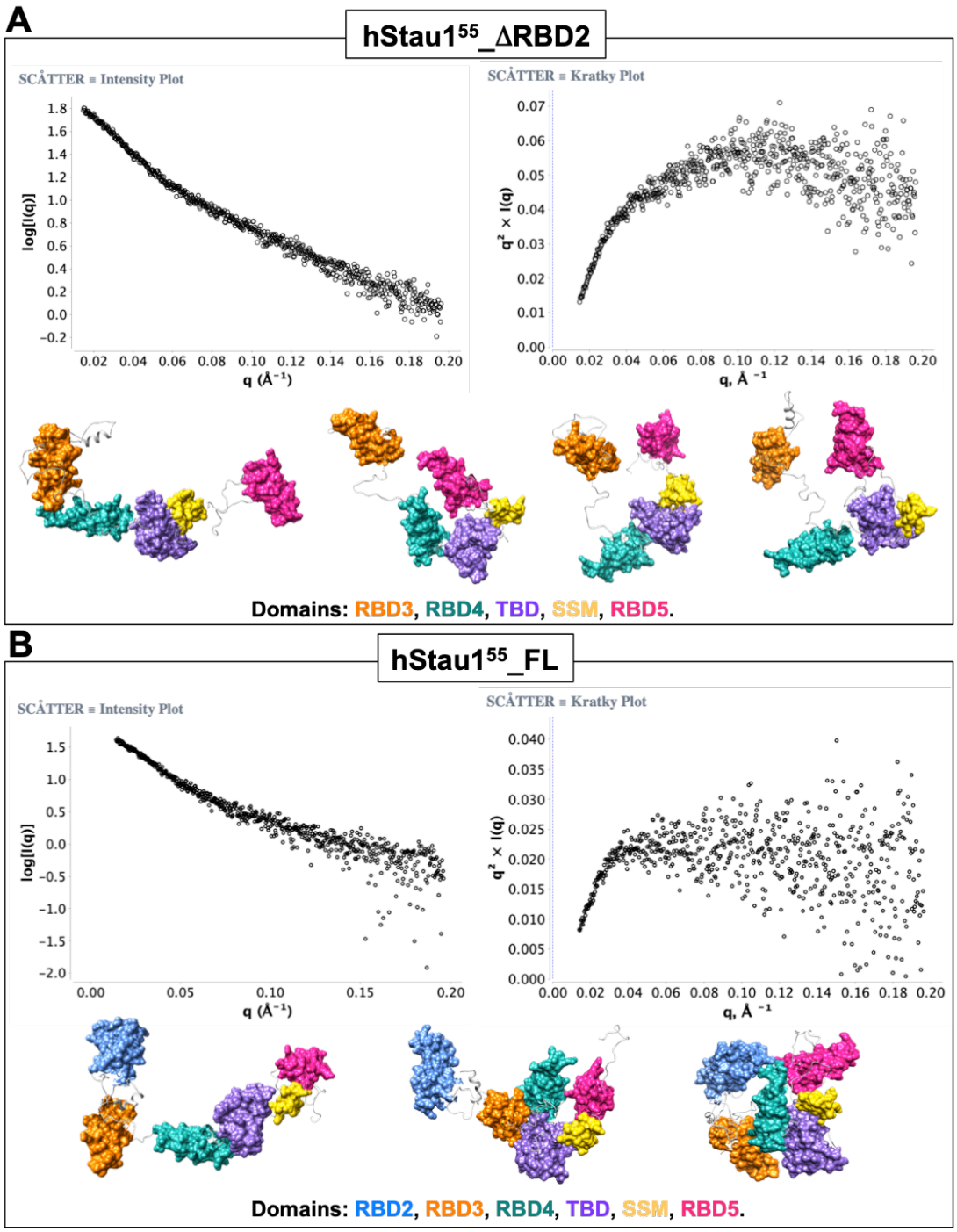


Figure 3.16: Small angle X-ray scattering (SAXS) of Staufen proteins. A: SAXS curve, Kratky analysis and EOM models generated for hStau1^{55}_ΔRBD2 [R_g *ensemble*= 50.26 Å, D_{max} *ensemble*= 166.96 Å, R_{sigma}= ~80.2% (~85.68%)/ X^2=2.30]. B: SAXS curve, Kratky analysis and EOM models generated for hStau1^{55}_FL [R_g *ensemble*= 48.11 Å, D_{max} *ensemble*= 155.2 Å, R_{sigma}= ~74.45% (~86.17%)]/ X^2=2.11].

The structural information gathered from the SAXS models for hStau1^{55}_FL and hStau1^{55}_ΔRBD2 were validated using 2D NMR (Figure 3.17).

^{15}N,^{1}H-TROSY spectra were recorded for individual and tandem domains, as well as for the full-length protein and its truncation mutant ΔRBD2. The large number of peaks in the central area of the spectra of both hStau1^{55}_FL and hStau1^{55}_ΔRBD2 indicates the presence of a large number of amino acids in disordered regions. This agrees well with the presence of long unstructured linkers that confer flexibility to the proteins.

Moreover, the similar lineshapes and lack of significant chemical shift perturbations between the spectra recorded for individual and tandem domains (Figures 3.3 and 3.4) show that any interactions between adjacent domains are very limited and that each of the domains constituting the tandem constructs tumbles independently in solution, thanks to the intervening linkers' flexibility. In addition, it is possible to reconstitute the spectra of hStau1^{55}_FL and hStau1^{55}_ΔRBD2 almost entirely by overlaying those obtained for individual domains, proving that the domains tumble independently in the full-length protein.

Together, our data suggest that hStau1^{55} is an extremely flexible protein and its domains can adopt several positions relative to each other, without inter-domain interactions. Thanks to the flexibility of the linkers, the protein adopts an elongated conformation in solution and its domains behave as beads on a string. Connector regions are crucial players in Staufen allostery and conformational changes are in line with recent studies on the role of dynamic linkers in the modulation of protein function (Papaleo et al., 2016).

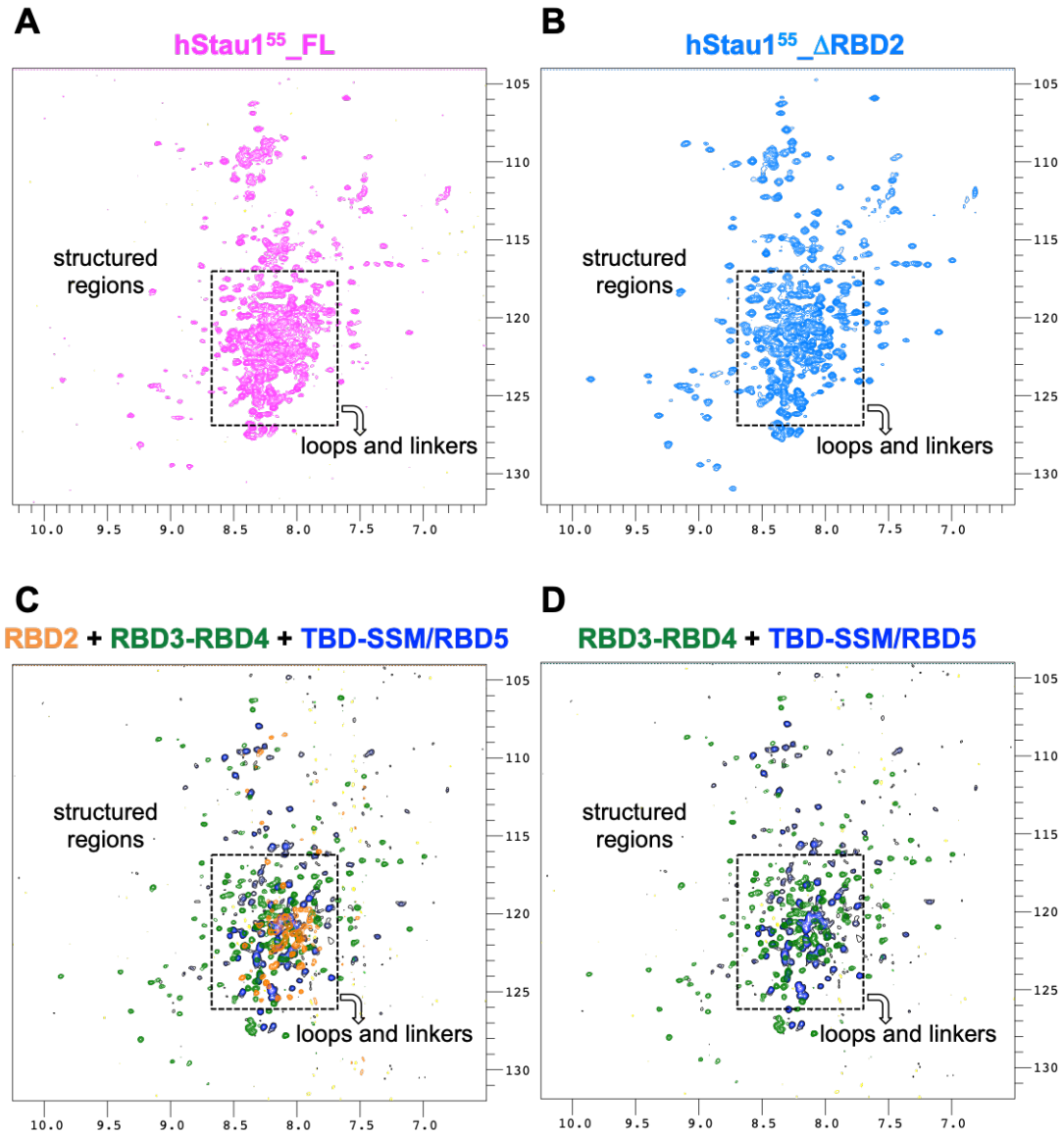


Figure 3.17: NMR of Staufen proteins. A: NMR spectra of hStau1^{55}_FL. B: NMR spectra of hStau1^{55}_ΔRBD2. C: Reconstitution of hStau1^{55}_FL spectra by overlapping of RBD2, RBD3-RBD4 and TBD-SSM/RBD5 NMR spectra. D: Reconstitution of hStau1^{55}_ΔRBD2 spectra by overlapping of RBD3-RBD4 and TBD-SSM/RBD5 NMR spectra.

3.3.1.4 Linker flexibility mediates RBD3 and RBD4 rearrangement

The fitting of a representative subset of SAXS models for individual or tandem domains in the SAXS models of hStau1^{55}_ΔRBD2 and hStau1^{55}_FL is shown in Figure 3.18 and 3.19, respectively. The SAXS data and models obtained for individual and tandem domains have been previously discussed in section 3.3.1.1 (Figures 3.5-3.13). A representative subset of solution scattering models of individual and tandem domains (Figure 3.18B) was chosen in order to interpret the domain rearrangement observed in the models corresponding to the different conformations that hStau1^{55}_ΔRBD2 adopts in solution (Figure 3.18A). From the fitting proposed in Figure 3.18C, it is possible to observe that in hStau1^{55}_ΔRBD2 the linker connecting RBD3 to RBD4 can be completely or partially distended. SAXS models obtained for the construct RBD3-RBD4 well describe the behaviour of these two domains also when they belong to hStau1^{55}_ΔRBD2, showing that the presence of TBD and SSM/RBD5 does not have a great impact on RBD3-RBD4 rearrangements.

On the contrary, the models obtained for the construct RBD3-RBD4 cannot be used for the interpretation of domains rearrangements in hStau1^{55}_FL (Figure 3.19A) and a different subset of solution scattering models (Figure 3.19B) needs to be used in the fitting to describe the conformational changes of the full-length protein (Figure 3.19C). In fact, the solution models obtained for the tandem domain RBD3-RBD4 in isolation (Figure 3.11) show that a long linker, which is extremely elongated, connects these two domains. On the other hand, our hStau1^{55}_FL models show the coexistence of three main conformations in solution for which it is interesting to notice the relative movement of RBD3 and RBD4 and their closer proximity, possibly in order to elicit the binding of RNA targets (Figure 3.20). In the most elongated model of the full-length protein (model 1), RBD3 and RBD4 are in an ‘open’ conformation that resembles the one adopted by the tandem domain on its own. However, in the other two models (models 2 and 3), RBD3 and RBD4 are ‘pulled’ towards each other by conformational changes of the connecting linker, interestingly resembling the recently deposited structure of the hStau1^{63} RBD3-RBD4 construct bound to Arf1 SBS (Lazzaretti et al., 2018) (represented in red and blue in Figure 3.20).

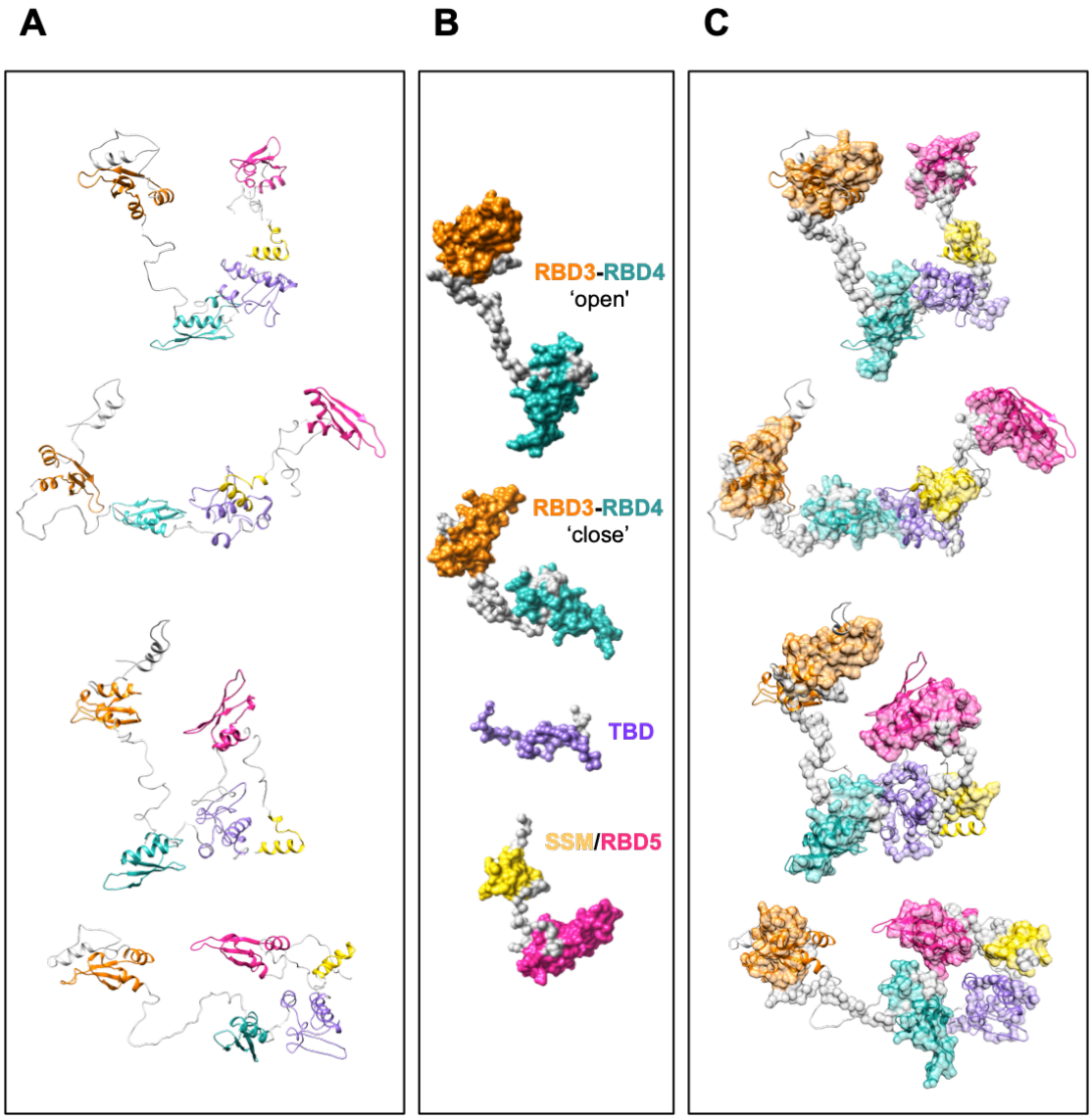


Figure 3.18: Fitting of SAXS models for Staufen1 domains in the SAXS models obtained for hStau1^{55}_ΔRBD2. A: EOM models generated for hStau1^{55}_ΔRBD2. B: Subset of selected domains and tandem domains models. C: Fitting of representative models for domains and tandem domains in the EOM models generated for hStau1^{55}_ΔRBD2.

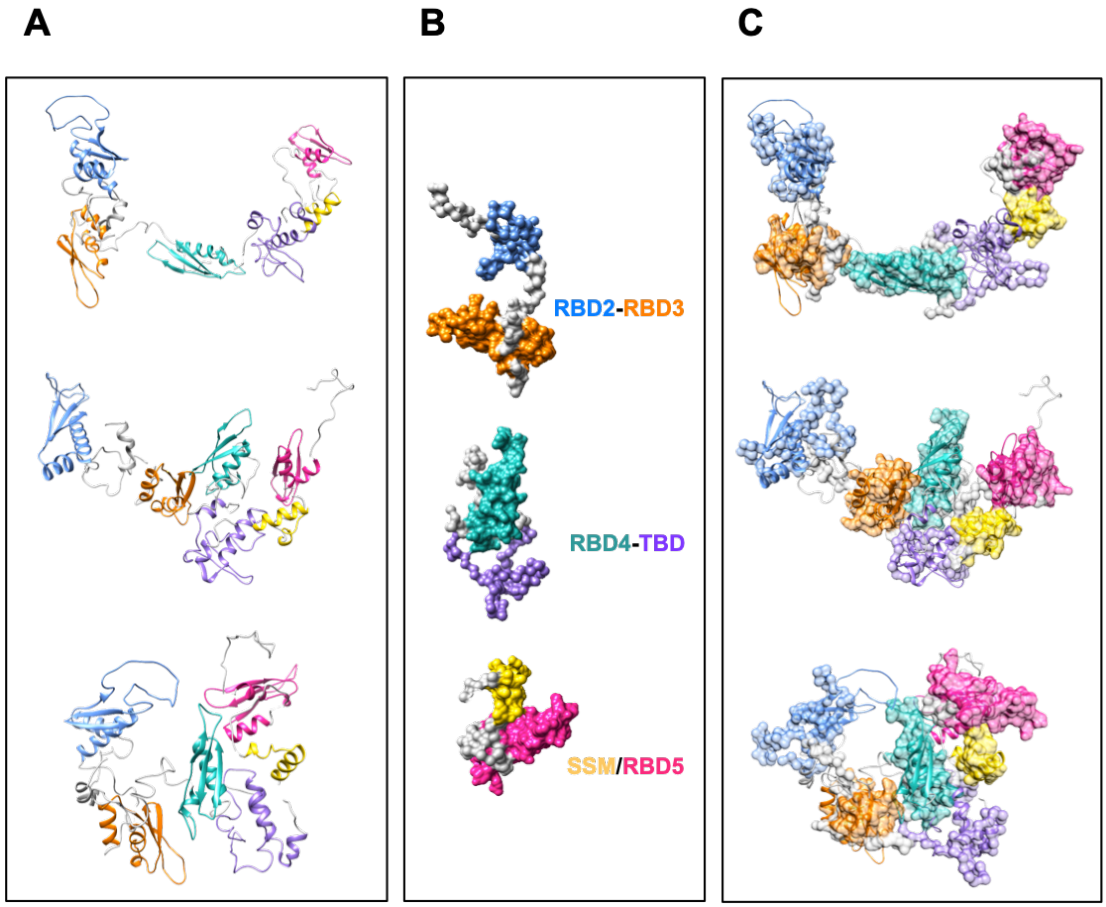


Figure 3.19: Fitting of SAXS models for Staufen1 domains in the SAXS models obtained for hStau1^{55}_FL. A: EOM models generated for hStau1^{55}_FL. B: Subset of selected domains and tandem domains models. C: Fitting of representative models for domains and tandem domains in the EOM models generated for hStau1^{55}_FL.

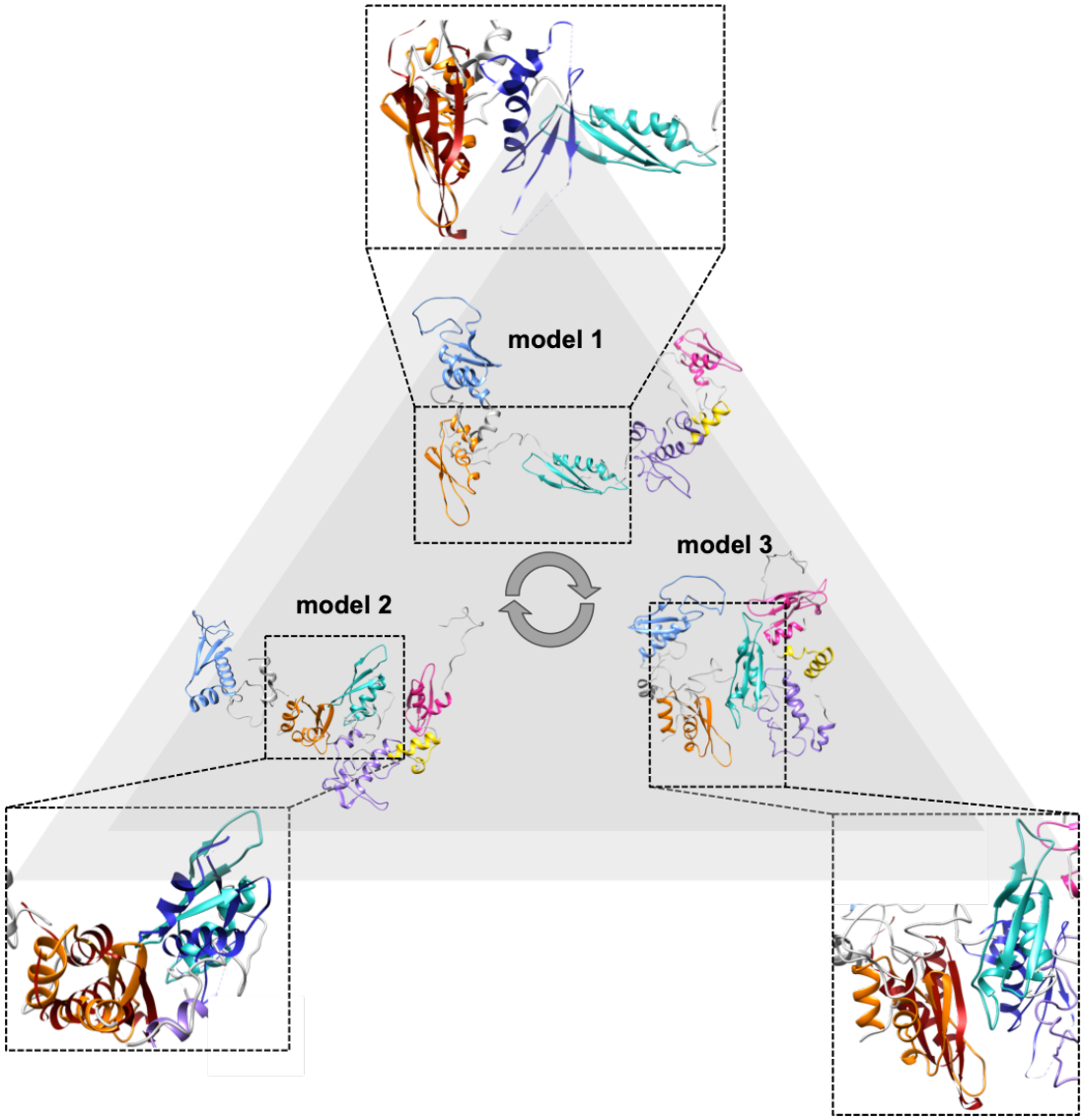


Figure 3.20: RBD3 and RBD4 rearrangements in hStau1^{55}_FL can explain its plasticity in the binding of diverse dsRNA targets. The SAXS models obtained for hStau1^{55}_FL show that RBD3 (orange) and RBD4 (cyan) transit from an 'open' to a 'more closed' conformation. A direct comparison of our models with the crystal structure of the hStau1^{63} RBD3-RBD4 construct bound to Arf1 SBS^{43} (RBD3 displayed in red and RBD 4 in blue) shows that RBD3 and RBD4 are 'pulled' towards each other by conformational changes of the connecting linker, possibly in order to elicit dsRNA binding.

Cryo-EM coupled to single particle analysis was exploited in order to gather more information about the architecture of Staufen protein and, in particular, about the role of RBD2 in triggering the domains rearrangement observed in the full-length models obtained from small angle scattering analysis.

The continuous advances in the electron microscopy field, and in particular the use of phase plate and direct electron detector, allowed us to directly visualise the monomeric hStau1^{55}_FL protein. Importantly, Staufen is just below the limit of size of protein that is possible to study using cryo-EM analysis (to date) and this represents a major difficulty in both microscopy and image processing stages. The data collection was carried out at the LMB-MRC (Cambridge, UK) by Dr Giuseppe Cannone and the data was processed by Dr Laura Spagnolo (Figure 3.21).

Nevertheless, the low-resolution cryo-EM density map obtained for hStau1^{55}_FL is in agreement with our SAXS studies and shows that Staufen is an elongated protein with dimensions 126x63x43 Å (Figure 3.22, top).

As shown in Figure 3.22, the protein can be described to have an L-shaped architecture, which can be divided into three main portions: a “head”, a “body” and a “foot”. Moreover, it is possible to observe a thin element between the “head” and the “body”, which can be described as a “neck”.

Even though the SAXS models obtained for hStau1^{55}_FL show that the protein could adopt at least three major different conformations in solution, cryo-EM analysis reveals only an extended architecture, which resembles the one represented by the most elongated SAXS model (hStau1^{55}_FL EOM model 1). In fact, the calculated SAXS curve of the DAMMIN model generated from the cryo-EM density map (using EM2DAM) is extremely similar to the curve calculated for the most extended hStau1^{55}_FL SAXS model (using CRYSOL), which represents the scattering contribution of this conformation to the total scattering intensity (Figure 3.22, bottom).

The discrepancy between the several conformations revealed using SAXS studies and the single architecture obtained with cryo-EM analysis, could be due to an effect of cryo-EM experimental condition during sample preparation (e.g. surface tension and freezing) or to the fact that more compact architectures are extremely difficult to distinguish from the background during particle picking, due to the very small size of the protein.

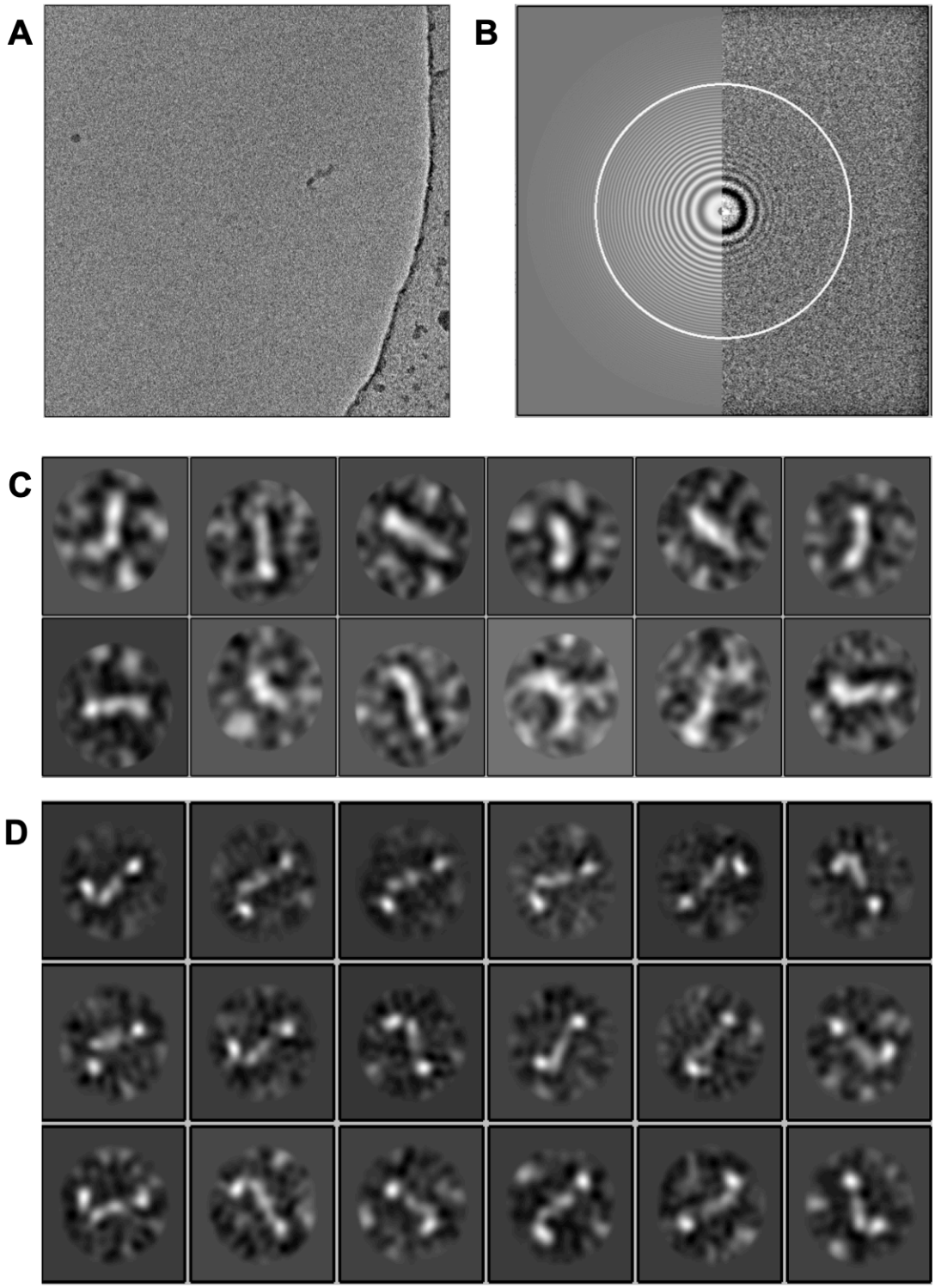


Figure 3.21: Cryo-EM single particle analysis of hStau1[55]. A: Representative micrograph. B: CTF for the representative micrograph. C: Representative 2D classes from 2245 manually picked single particles, boxsize = 208 Å. D: Representative 2D classes from 198,088 autopicked single particles.

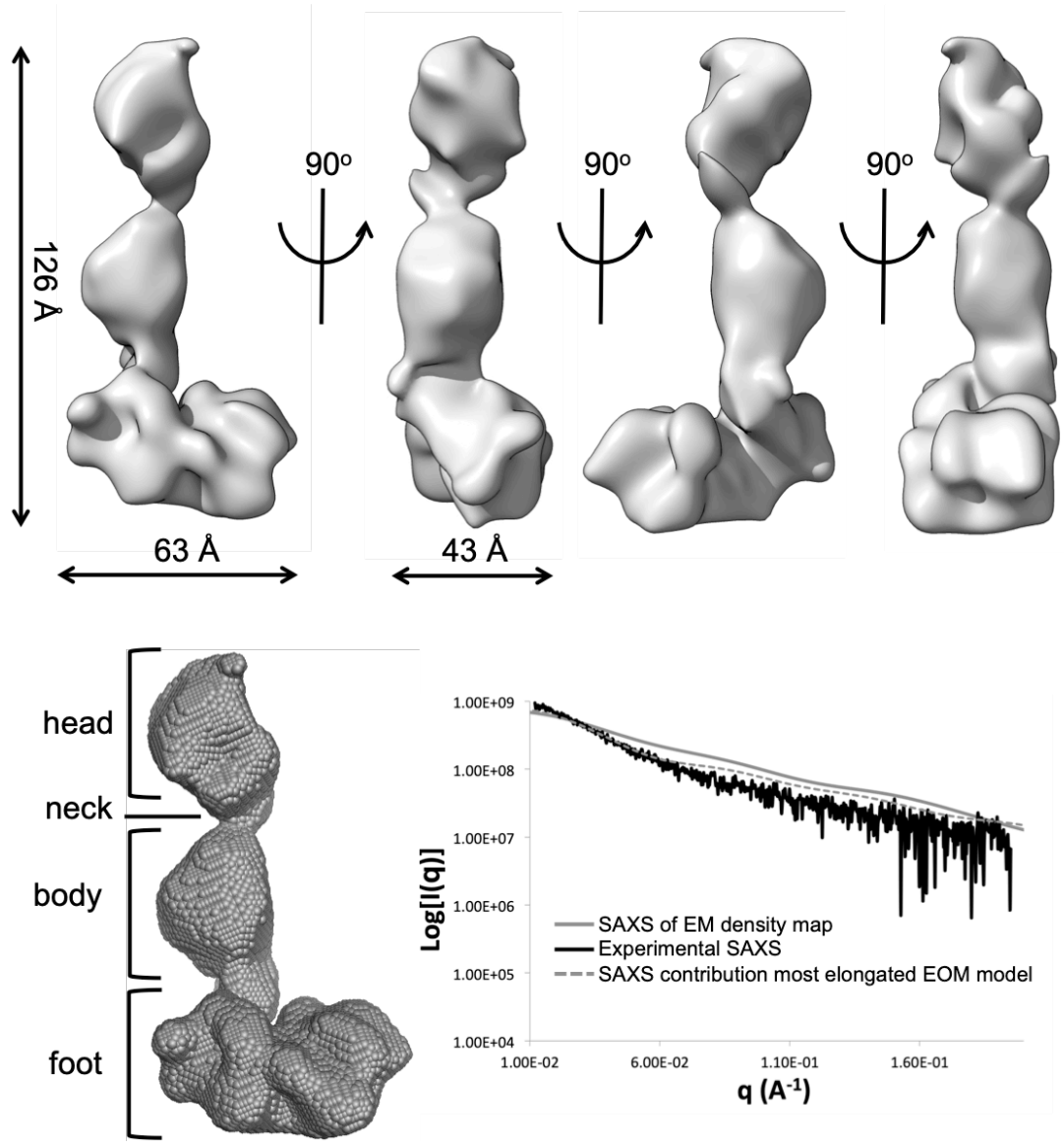


Figure 3.22: Architecture of hStau1[55] full-length protein. Top: Cryo-EM structure of full-length Staufen1. Bottom left: DAMMIN model calculated on the EM density map (EM2DAM). Bottom right: Theoretical SAXS curve calculated for the cryo-EM map (grey, continuous) overlaid to the experimental SAXS curve for the ensemble of conformations (black) and to the SAXS curve calculated for the most elongated hStau1[55] EOM model (grey, dashed).

The domain arrangement observed in the cryo-EM density map is consistent with the beads-on-a-string model, as supported also by NMR analysis.

SAXS studies of hStau1^{55}_FL show that RBD2 and RBD3 can be in close proximity to each other. Therefore, we fitted the more compact SAXS model obtained for the tandem domain RBD2-RBD3 (Figure 3.10) at the foot of the cryo-EM map. The remaining density allows space for the other domains (RBD4, TBD, SSM and RBD5) to be fitted within the map (Figure 3.23B). As previously described, the solution models obtained for the tandem domain RBD3-RBD4 (Figure 3.11) show that a long linker, which is extremely elongated, connects these two domains. However, in our hStau1^{55}_FL SAXS models, RBD3 and RBD4 are in closer proximity. In the cryo-EM density map, the electron density corresponding to RBD4 is fitted according to the spatial domains organization of the more compact SAXS model obtained for the tandem domain RBD4-TBD (Figure 3.12). It follows that RBD4 must undergo a spatial reorganization, as described in Figure 3.23A. Moreover, since Staufen dimerization is mediated by the domain swapping motif SSM, the arrangement proposed by the fitting of one of the SAXS models obtained for SSM/RBD5 (Figure 3.9) within the “neck” and “head” of the density map, respectively, would allow the SSM of each monomer to be accessible for homo-dimerization.

Taken together, these data show that not only the presence of RBD2 triggers a spatial reorganization of RBD3 and RBD4, which is indeed mediated by the linker between these last two domains, but also the importance of RBD3-RBD4 relative position and rearrangements on dsRNA binding. Therefore, we propose that these two domains can change mutual orientation depending on the structure of the RNA target, in order to effectively bind different substrates in distinct biological contexts. Moreover, dsRBDs that are not involved in the binding of RNAs, such as RBD2 and RBD5, can adopt multiple conformations in the full-length protein, not only to elicit protein dimerization, but also to regulate Staufen structural plasticity and multi functionality *in vivo*. All in all, our solution studies demonstrate that Staufen protein can adopt several conformations thanks to long linkers that facilitate domains rearrangement, providing a clue on the structural background for its role in multiple biological pathways.

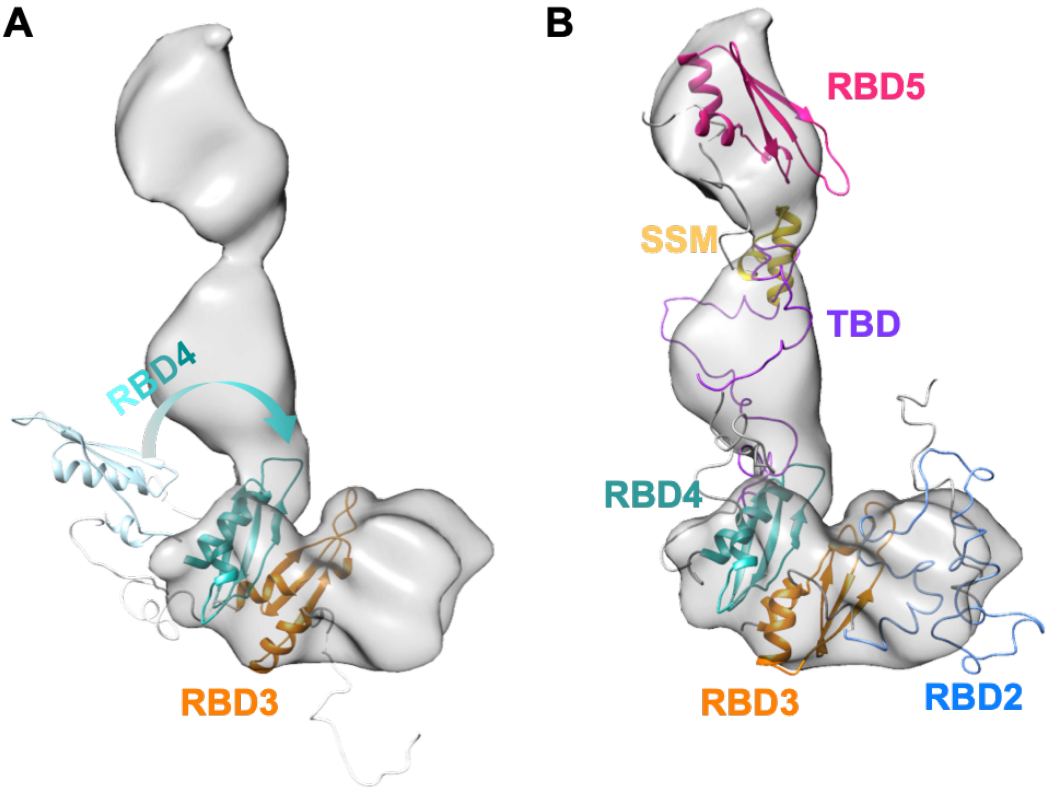


Figure 3.23: Fitting of SAXS models for Staufen1 domains in the cryo-EM density map of the full-length protein. A: Fitting of representative models for the ensemble of the SAXS models generated by EOM for the tandem domains RBD3-RBD4 (orange and turquoise) and RBD4-TBD (for clarity, only RBD4 is shown in teal) showing RBD4 rearrangement in the full-length protein. B: Fitting in the EM density map of representative models for the ensemble of the SAXS models generated by EOM for the tandem domains RBD2-RBD3 (blue and orange), RBD4-TBD (teal and purple) and SSM/RBD5 (yellow and magenta).

3.3.2 Staufen/Arf1 RNA interaction

3.3.2.1 RBD2 is important in the stabilization of Staufen/Arf1 complex

Previous studies have shown that Staufen-binding sites (SBSs) are located in intra-molecular double-stranded regions, typically between 9 and 12 bp, at the 5'UTR or 3'UTR of mRNAs and that Staufen binding mediates their translation or decay, respectively (Dugre-Brisson et al., 2005, Kim et al., 2005). Moreover, SBSs can be located in inter-molecular double-stranded regions formed between mRNAs and long noncoding RNAs *via* Alu element base-pairing (Gong and Maquat, 2011).

Human ADP-Ribosylation Factor 1 (ARF1) mRNA represents the best characterized Staufen target to date, together with *D.melanogaster* oskar (osk) and bicoid (bcd) mRNAs (Ferrandon et al., 1997, Irion et al., 2006, Micklem et al., 2000). The 3'UTR of Arf1 (~300 bp) forms an apex structure and mutagenesis analysis showed that the 19 bp stem is required for Staufen binding (Kim et al, 2007) (Figure 3.24).

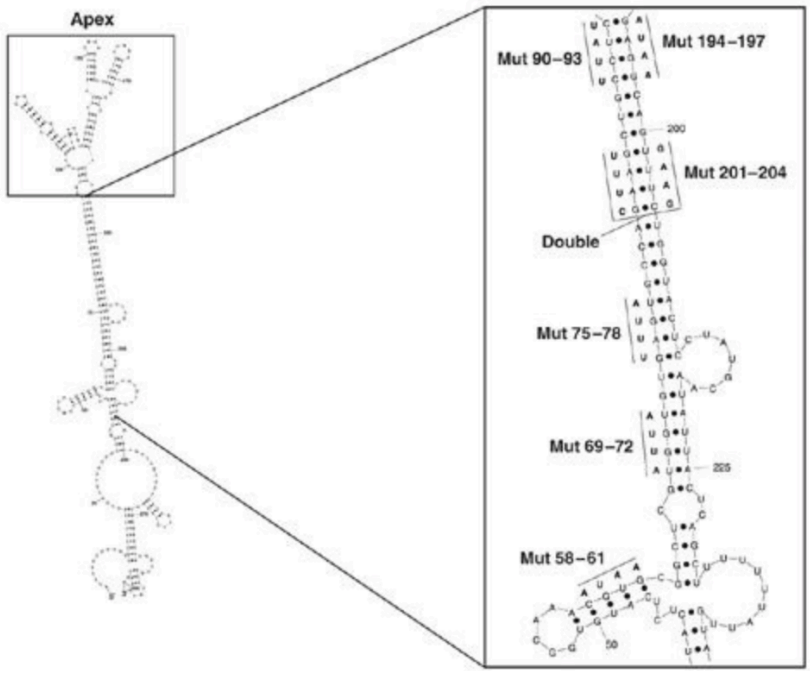
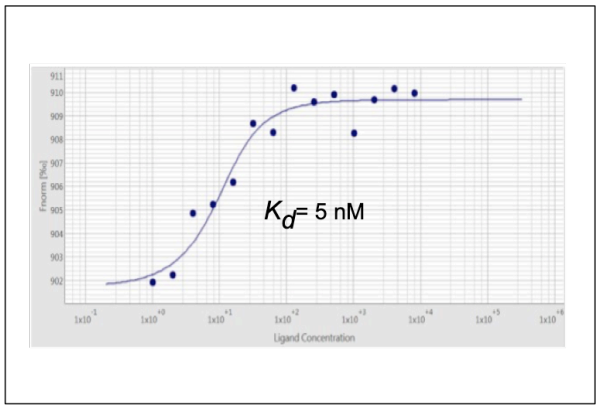


Figure 3.24: Structure prediction of ARF1 3'UTR. Model of the secondary structure of ARF1 showing nucleotides 1 through 300 and mutagenesis studies, which confirmed the 19bp stem to be the major SBS. The plot was generated using Sfold v2.0 software (Wadsworth Bioinformatics Center) by Kim *et al*, 2007. (Adapted from [Kim *et al*, 2007])

Consequently, the Arf1 mRNA bound to Staufen is targeted for degradation through Staufen-mediated decay. On the other hand, the 19 bp stem is not the only Staufen-binding site within the 3'UTR of Arf1. In fact, two additional shorter SBSs have been identified, by high-throughput studies, in shorter double-stranded regions within the apex (Sugimoto et al., 2015). Lazzaretti *et al*, using fluorescence anisotropy and X-ray crystallography, recently characterized the interaction between hStau1^{63} proteins and the 19 bp stem of Arf1 mRNA. The crystal structure, at 2.9 Å resolution, shows that the tandem RBD3-RBD4 of hStau1^{63} recognizes A-form RNA, mainly through electrostatic interactions with the RNA backbone (Lazzaretti et al., 2018).

In our study, the interaction between hStau1^{55} proteins and the entire 3'UTR of its RNA target Arf1 was investigated with microscale thermophoresis (MST) analysis, during a demonstration of the Monolith NT.115Pico instrument by NanoTemper. As represented in Figure 3.25, MST analysis shows that hStau1^{55}_FL/Arf1 complex (K_d = 5 nM) is more stable than hStau1^{55}_ΔRBD2/Arf1 complex (K_d = 28 μM), even though the latter is probably under-estimated in this analysis as the reaction has not reached saturation. Our data on the interaction between hStau1^{55}_FL protein and Arf1 RNA are in agreement with the fluorescence anisotropy measurements by Lazzaretti *et al*, which show that purified recombinant hStau1^{63}_FL interacts with the 19 bp stem of Arf1 with high affinity (6 ± 1 nM). Moreover, an N-terminal deletion of hStau1^{63}, removing RBD2, only minimally affects the RNA-binding affinity (9 ± 3 nM) (Lazzaretti et al., 2018). On the other hand, our MST data show that RBD2 deletion in hStau1^{55} dramatically reduces the affinity of the protein to Arf1 RNA (K_d: from 5 nM to 28 μM). However, the fact that we used the entire 3'UTR of Arf1, instead of the 19 bp stem only, might be the basis of this apparent discrepancy. In fact, several Staufen monomers could bind at three different sites in the 3'UTR of Arf1, and this interaction could be stabilized by RBD2 via protein-protein interactions, as this domain is not involved in the direct binding of RNA.

hStau1_FL/Arf1 complex



hStau1_ΔRBD2/Arf1 complex

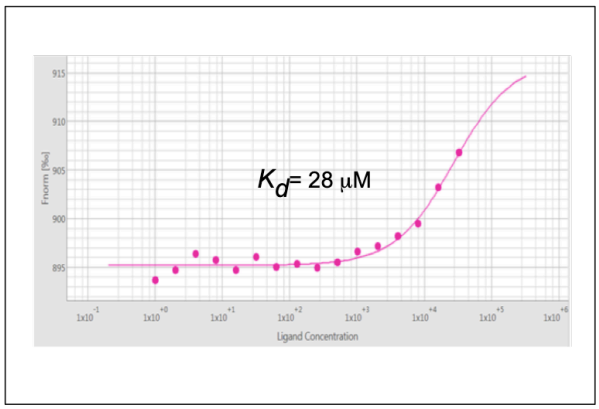


Figure 3.25: MST assay to study the interaction of Staufen proteins with Arf1 RNA. Top: Titration curve of NT-647 labeled hStau1^{55}_FL protein (10 nM) and Arf1 RNA (33 µM-1 nM), previously incubated for 30 minutes at room temperature. Bottom: Titration curve of NT-647 labeled hStau1^{55}_ΔRBD2 protein (10 nM) and Arf1 RNA (33 µM-1 nM), previously incubated for 30 minutes at room temperature. MST analysis was performed in premium glass capillaries using the Monolith NT.115Pico instrument.

3.3.2.2 Negative staining of hStau1^{55}_ΔRBD2 /Arf1 complex

Negative staining coupled to single particle analysis was employed in order to gather more information about the architecture of Staufen protein in complex with its RNA target Arf1.

Data collection was carried out for both hStau1^{55}_FL/Arf1 and for hStau1^{55}_ΔRBD2/Arf1 complexes purified from glycerol gradient. However, the higher percentage of glycerol (~45-50%) in the sample of the full-length protein/RNA complex interfered with signal-to-noise ratio and no further image processing analysis were undertaken.

From the micrographs of the hStau1^{55}_ΔRBD2/Arf1 complex dataset (Figure 3.26A), ~290,000 single particles (SP) were auto-picked with EMAN2, using 250 Å masks. The single particle images were extracted into boxes (128x128 pixels) and band-pass filtered (Figure 3.26B). The dataset was subjected to sequential rounds of alignments and 2D classification in RELION, using a 200 Å mask, in order to polish the dataset and improve the resulting class-average images (Figure 3.26C). ~43,000 particles were then sorted with RELION into three 3D classed (Figure 3.26D) using as reference an *ab initio* initial model generated with EMAN2.

As shown in Figure 3.26D, the three 3D classes generated for hStau1^{55}_ΔRBD2/Arf1 are fairly dissimilar and this could be due to partial assembly and/or dissociation of the complex. This heterogeneity can also be observed in the 2D classification (Figure 3.26C), as the class averages show smaller and bigger assemblies. Moreover, the stoichiometry of the complex is not known, and this limits the possibility of further interpretation of the low-resolution electron density map obtained from negative staining.

Nevertheless, the 3D class “class 1” shows some common features with one of our SAXS model for the monomeric hStau1^{55}_ΔRBD2 (Figure 3.27). In fact, it is possible to notice that the protein/RNA complex shows some “triangular” densities, which could be assigned to RNA-binding domains. Moreover, the volume of the density map (143.6 Å^3) suggests that at least two hStau1^{55}_ΔRBD2 monomers (41.08 Å^3) might be bound to Arf1 RNA and that the structured RNA apex could also contribute to the total electron density.

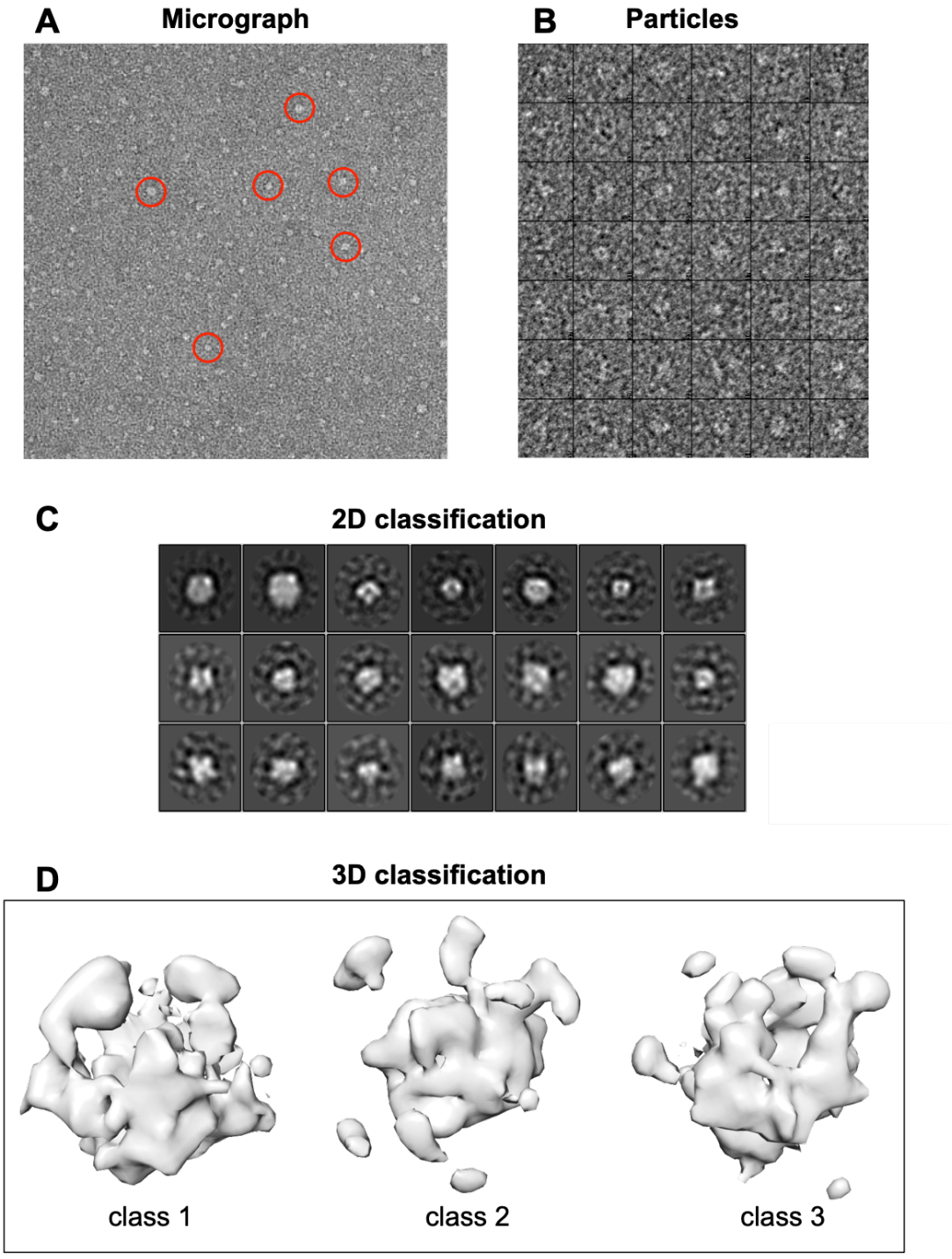


Figure 3.26: Single particle analysis of negative stained grids of hStau1^{55}_ΔRBD2 protein and Arf1 RNA complex purified from glycerol gradient. A: Characteristic micrograph of hStau1^{55}_ΔRBD2 protein and Arf1 RNA complex, red circles are used to show single particles. B: Subset of individual particles extracted from auto-picking (128x128 pixels, 1.9 Å/pixel). C: Class averages obtained from 2D classification using a 200 Å mask. D: 3D classes obtained from 3D classification using a 200 Å mask.

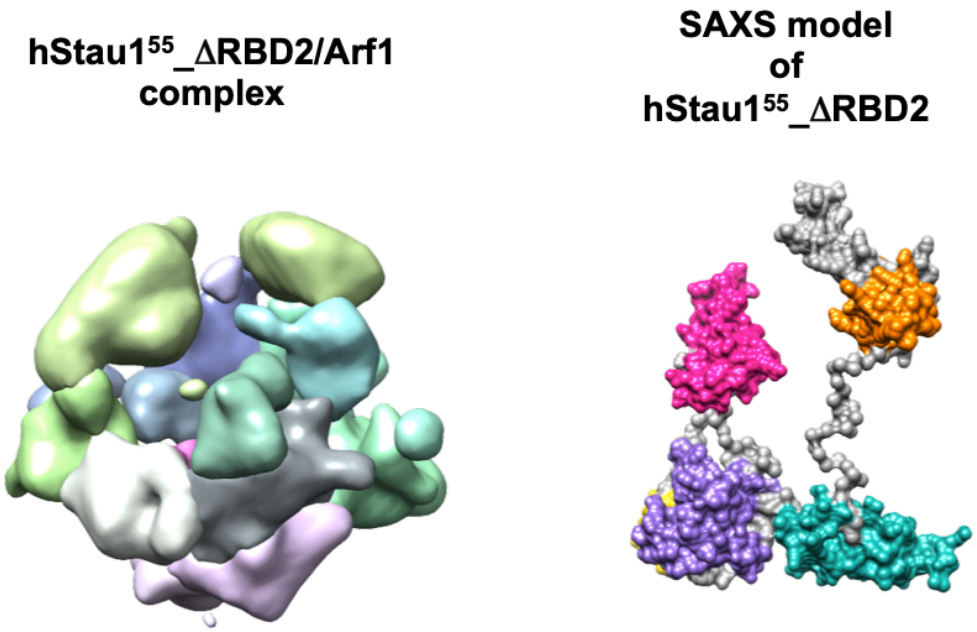


Figure 3.27: Segmented map of hStau1^{55}_ΔRBD2/Arf1 complex and representative SAXS model of hStau1^{55}_ΔRBD2. Left: Segmented electron-density map of 3D class n.1 of hStau1^{55}_ΔRBD2 protein and Arf1 RNA complex obtained with single particle analysis coupled to negative staining electron microscopy. Right: Representative SAXS model of hStau1^{55}_ΔRBD2 obtained with EOM showing characteristic domains organization of monomeric protein (RBD3 in orange, RBD4 in teal, TBD in purple, SSM in yellow, RBD5 in magenta and connecting linkers in grey).

3.3.3 hUpf1 protein

3.3.3.1 Overexpression and purification of hUpf1 protein

E. coli Rosetta™(DE3) pLysS cells were transformed by heat-shock method with pTWO-E_Strep-tag vector. To test overexpression conditions for the production of recombinant hUpf1, cells were treated with different concentrations of IPTG (0.25 mM and 0.5 mM) and aliquots were taken at different time-points (2,4 hours and o/n). The overexpression was carried out at 23°C, in order to allow proper folding of this relatively big (~125 kDa) human protein in *E. coli*. Protein production was assessed by SDS-PAGE and Western Blot analysis (Figure 3.28A). Protein overexpression was induced by adding 0.25 mM IPTG and culturing the cells at 23°C for 4 hours. In fact, after o/n induction the amount of protein was extremely low, likely because of protein degradation.

Four litres of culture were grown in order to obtain enough soluble protein. Cells were lysed by sonication and soluble protein extracts were separated from the cell pellets by centrifugation. Strep-hUpf1-6His protein was firstly purified by nickel chromatography in high-salt buffer in order to remove nucleic acid contamination. Elution was performed with an imidazole gradient and the fractions containing protein were analysed by 8% SDS-PAGE analysis (Figure 3.28B). Subsequently, the fractions containing STREP-hUpf1-6His were pooled and incubated with streptactin resin, hUpf1 full-length protein was eluted by addition of desthiobiotin and analysed in an 8% SDS-PAGE gel (Figure 3.28C). The elution from the streptactin resin was then subjected to size exclusion chromatography (SEC) using a Superdex 200 Increase 10/300 GL column. The SEC trace and SDS-PAGE analysis of the elutions are shown in Figure 3.28D. The yield of purified hUpf1 full-length protein was extremely low and further studies were limited to biochemical assays in order to investigate the interaction between Staufen proteins, Arf1 RNA and hUpf1 protein.

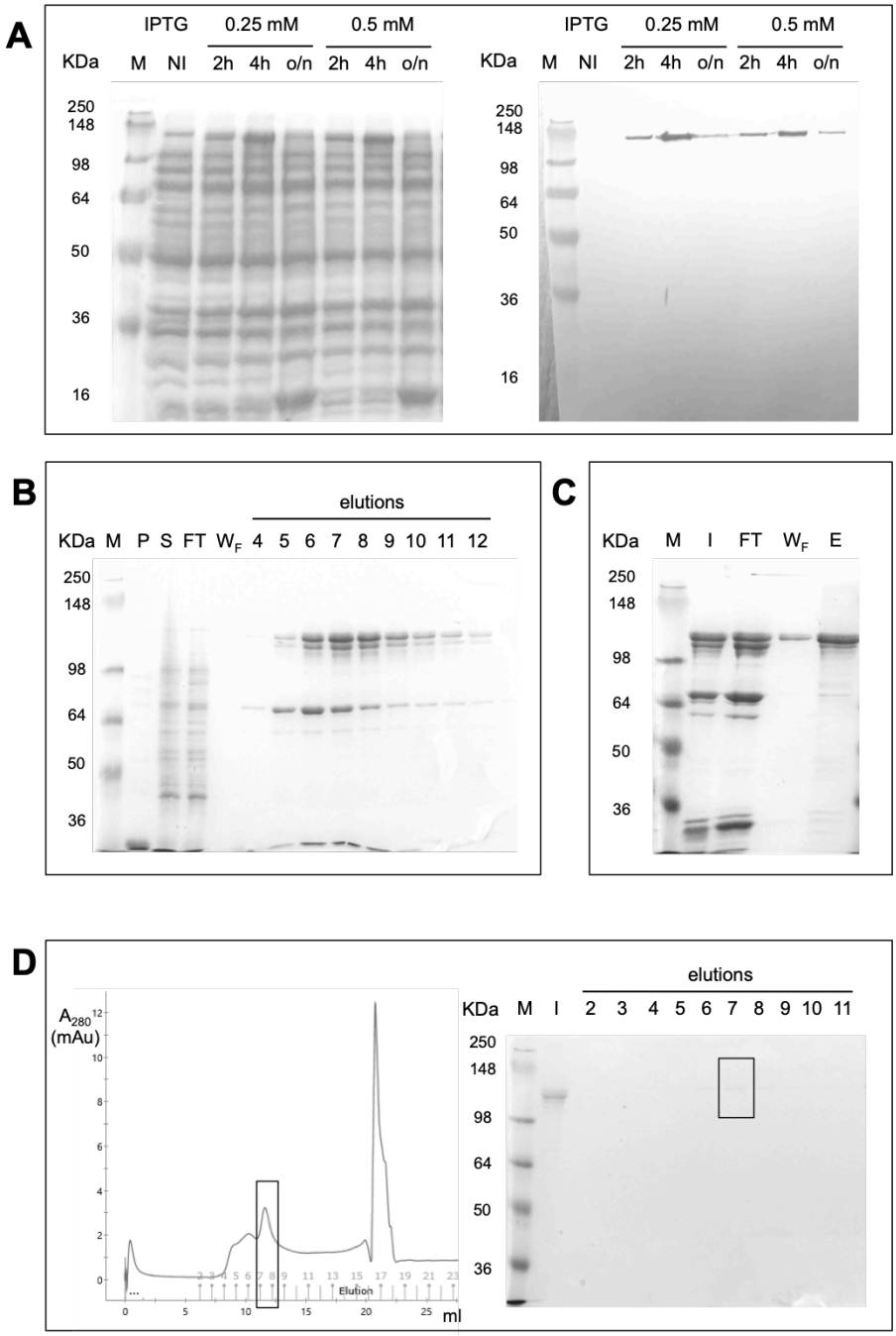


Figure 3.28: Overexpression and purification of hUpf1. A: SimplyBlue SafeStain staining of 10% SDS-PAGE gel of overexpression test in *E. coli* Rosetta™(DE3) pLysS cells grown in TB and induced at 23°C with 0.25 mM and 0.5 mM IPTG for 2,4 hours and o/n (left). Western Blot analysis of 10% SDS-PAGE gel of overexpression test: the nitrocellulose membrane was incubated o/n at 4°C with anti-His antibody conjugated with alkaline phosphatase (1:4000) and developed using the SIGMA-FAST™ BCIP/NBT reagent (right). B: SimplyBlue SafeStain staining of 8% SDS-PAGE gel of nickel chromatography purification step in buffer B (25 mM HEPES pH 7.4, 1 M NaCl, 1 mM $MgCl_2$, 1 µM $ZnCl_2$, 500 mM imidazole). C: SimplyBlue SafeStain staining of 8% SDS-PAGE gel of streptactin purification step in buffer B supplemented with 2.5 mM desthiobiotin. D: SEC trace of hUpf1 full-length purified in buffer C (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$) using a Superdex 200 Increase 10/300 GL SEC column at 0.7 ml/min (left) and SimplyBlue SafeStain staining of 8% SDS-PAGE gel of SEC purification step (right).

3.3.3.2 Study of the interaction between Staufen proteins, Arf1 RNA and hUpf1 protein

Previous *in vivo* studies showed that Stau1 binds directly to Upf1 and elicits mRNA decay when bound downstream of a termination codon, *via* SMD pathway instead of the conventional NMD, which would require Upf2 and Upf3 to recruit Upf1 and consequently to elicit RNA decay (Kim et al., 2005).

The interaction between hStau1^{55}_FL or hStau1^{55}_ΔRBD2 protein, Arf1 RNA and hUpf1 protein was studied by incubating an equimolar concentration of each component. Subsequently, complex formation was evaluated by BN-PAGE analysis (Figure 3.29).

While an extra band, likely to correspond to the protein/RNA complex, can be observed when hStau1^{55}_ΔRBD2 protein is incubated with Arf1 RNA, it is not possible to detect any difference in the bands pattern of hStau1^{55}_FL when in absence or presence of RNA. In fact, the full-length protein shows several bands, corresponding to monomer, dimer and higher order assemblies, which are particularly smeary, indicating the presence of different conformations. This makes it difficult to detect whether a band corresponding to hStau1^{55}_FL/Arf1 complex is present.

Moreover, the incubation of hUpf1 and Arf1 results in the appearance of an extra band in the gel, suggesting that purified recombinant hUpf1 is active and can bind RNA. However, the band corresponding to hUpf1/Arf1 complex is also present when Staufen proteins were in the reaction mix. This suggests that, in the Staufen/Arf1 mix, free RNA is present in solution and it is then bound by hUpf1.

Furthermore, using this technique, it is not possible to observe whether or not a Staufen/Arf1/hUpf1 complex is assembled and further experiments should be used to address this still open question.

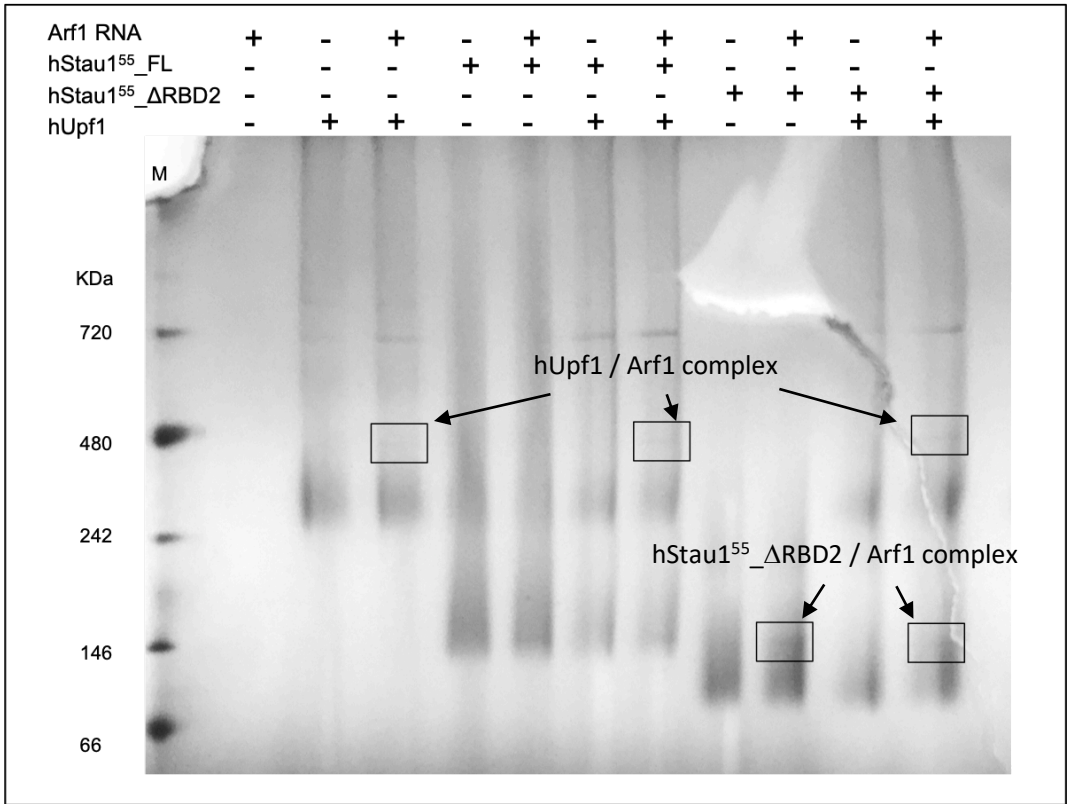


Figure 3.29: BN-PAGE analysis of Staufen proteins, Arf1 RNA and hUpf1 protein. 70 µM of each component were incubated at room temperature for 90 minutes in a final volume of 21 µl in buffer C (25 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$). 7.5 µl of each reaction were loaded in a 4-16% Bis-Tris NativePAGE and BN-PAGE was run for 1 hour at 150 V. Gel was de-stained o/n in 30% ethanol and 10% acetic acid.

3.4 Conclusions

Canonical dsRBDs are composed of an α1-β1-β2-β3-α2 secondary structure that folds in three dimensions to recognize dsRNA. Recently, structural and functional studies of divergent dsRBDs revealed adaptations that include intra- and/or intermolecular protein interactions, sometimes in the absence of detectable dsRNA-binding ability (Gleghorn and Maquat, 2014, Krovat and Jantsch). The number of canonical dsRBDs per polypeptide is highly variable in dsRBD-containing proteins (Barraud and Allain, 2012, Thomas and Beal, 2017, Macrae et al., 2006), ranging from one to five. There is no clear correlation between the number of dsRBDs and dsRNA binding affinity. Moreover, the extent of the contribution of specific dsRBDs to RNA substrate specificity is still an open question. Structural information reported on dsRBPs carrying one or more canonical dsRBD, such as Dicer (Macrae et al., 2006) and ADARs (Thomas and Beal, 2017, Barraud and Allain, 2012) highlights the importance of neighbouring domains for substrate specificity and enzymatic activity. Conformational flexibility of Dicer protein is also proposed to play a central role in dsRNA recognition and processing (Macrae et al., 2006) and this could be extended to other RBD-containing proteins, such as Staufen. The lack of structural information on this protein, for which to date we only had three-dimensional models of Staufen truncation mutants, in isolation or in complex with short RNA sequences or truncated protein interactors (Gleghorn et al., 2013, Ramos et al., 2000a, Ramos et al., 2000b, Jia et al., 2015, Lazzaretti et al., 2018) made its functional understanding particularly challenging. Here, we provided for the first time structural information on the full-length Staufen1 protein, using an integrated structural biology approach. Combination of hydrodynamic methods, homology modelling, small angle X-ray scattering, NMR and cryo-EM allowed us to show that Stau1 is a highly flexible protein, that recoils in solution from an elongated to a compact conformation in which the domains are in closer proximity but not interacting with each other. In this perspective, we propose that the flexible inter-domain loops possess a regulatory role in Staufen activity, allowing a high degree of freedom for recognition and binding of diverse RNA and protein targets and for the subsequent involvement of Staufen in very diverse aspects of RNA metabolism and regulation. As far as the dimerization of Stau1 is concerned, we

show that this is dependent on the presence of RBD2 and on protein concentration. Consistently with what was shown crystallographically, the SSM-RBD5 construct dimerized in solution. SAXS experiments show that the full-length protein adopts at least three main conformations in solution, which therefore can explain its ability to bind diverse RNA targets and protein partners. Our data provide the first structural insight into the mechanism adopted by Stau1[55] protein to elicit sometimes contrasting biological functions. Staufen seems to be able to re-arrange its domains thanks to long and disordered linkers, which confer plasticity to the protein, probably in order to increase the number of different mRNA targets and protein partners it can bind to. We can therefore think of Staufen as a "Swiss Knife" multi-tool. Furthermore, our preliminary negative staining EM studies provide the first structural information on a near-full-length Staufen protein (ΔRBD2) in complex with the 3'UTR of Arf1 RNA, which suggests that indeed Staufen undergoes domain rearrangements in order to elicit RNA binding, as suggested by our solution studies (Figure 3.30).

However, the question about whether or not Staufen/Arf1 complex is directly bound by hUpf1 in order to elicit mRNA decay, still remains unanswered and further experiments should be used to address this matter.

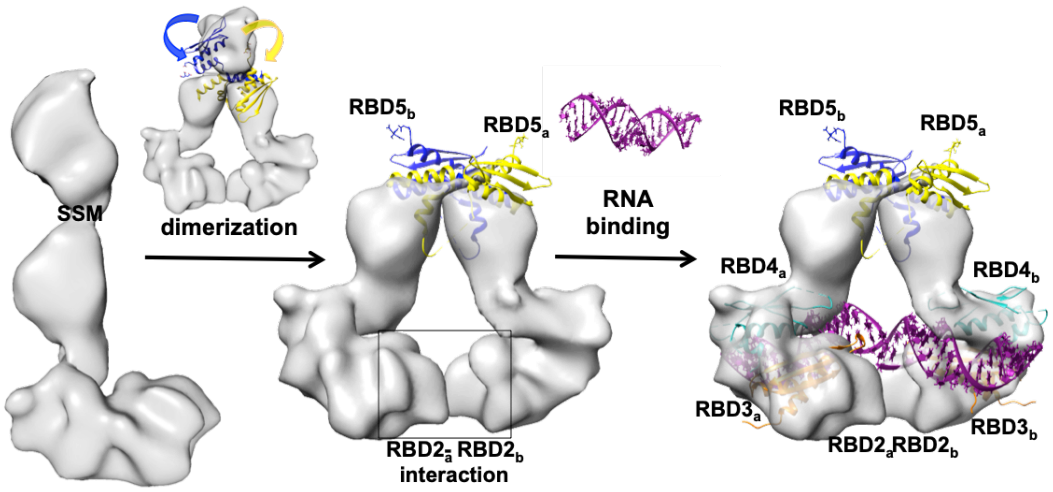


Figure 3.30: Model for the dimerization of Staufen1 protein and dsRNA recognition. Two Staufen1 monomers assemble through the SSMs and the resulting dimer is stabilised by $RBD2_a$-$RBD2_b$ interactions. $RBD5_a$ and $RBD5_b$ undertake a spatial rearrangement from the position they assume in the monomer (EM density map) to the position they assume in the crystallographic structure of the dimer SSM-RBD5 (Gleghorn et al., 2013) (PDB ID: 4DKK). In each monomer, RNA binding is mediated by RBD3 and stabilised by RBD4 (Lazzaretti et al., 2018) (PDB ID: 6HTU). Each RBD3 and RBD4 could bind two individual molecules of dsRNA or one dsRNA at two different sites.

Chapter 4

P. abyssi Mini-chromosome Maintenance protein

4.1 Introduction

Precise duplication of genetic material is required for the correct progression of the cell cycle and for cell division. Multiple proteins act together to separate the parental DNA and synthetize two new strands of complementary DNA. The set of proteins, whose concerted activities are necessary for a complete and high-fidelity genome duplication, is known as replisome.

Eukaryotes and Archaea employ a tightly regulated series of events to ensure correct DNA duplication, which must occur only once per cell cycle. The high degree of similarity between eukaryotic and archaeal proteins, involved in the replisome formation, allowed to study the archaeal DNA replication system as a model to provide a deeper understanding of the eukaryotic replication machinery.

In Eukaryotes, at the beginning of S-phase, the origin recognition complex (ORC), Cdc6, Cdt1 and MCM helicase come together to form the pre-replication complex (pre-RC), which initiates DNA melting at the origins of replication spaced along the chromosomes. After Cdc45 and GINS are incorporated into an intermediate pre-initiation complex (pre-IC) (Zou et al., 1997), the CDK and DDK protein kinases activate the Cdc45-MCM-GINS (CMG) complex (Aparicio et al., 2006, Aparicio et al., 2009, Costa et al., 2011, Ilves et al., 2010). Once activated, the two CMG complexes separate and the resulting replication forks travel in opposite directions unwinding genomic dsDNA to expose ssDNA template.

Mini-chromosome maintenance proteins (MCM) are AAA+ (ATPases associated with diverse cellular activities) superfamily members, which are known primarily as DNA helicases (Tye, 1999). However, their function is not limited to unwinding DNA and they have also been shown to be involved in origin melting

(Costa and Onesti, 2008), genome repair (Bailis and Forsburg, 2004, Bailis et al., 2008) and transcriptional regulation (Hubbi et al., 2011).

MCMs are replicative helicases highly conserved between Eukaryotes and Archaea; these proteins were first identified in yeast when mutations in their genes resolved in defective mini-chromosome maintenance, hence their name (Maine et al., 1984). Archaeal MCMs can be used as a simplified powerful tool for deciphering fundamental structural and functional features of the eukaryotic MCM complex since the MCM complexes of many Archaea form homohexamers from a single gene product, while eukaryotic MCM2-7 is a heterocomplex made of different polypeptide chains (Miller and Enemark, 2015, Brewster and Chen, 2010).

MCM proteins contains three domains, as shown in Figure 4.1A: a N-terminal domain, an AAA+ domain, and a C-terminal domain. The N-terminal domain is known to play roles in protein oligomerization, DNA binding, and the processivity of the helicase. The AAA+ domain is responsible for ATPase function and DNA unwinding. The small domain at the C-terminus may play a role in DNA-stimulated ATPase activity.

From the structural point of view, the N-terminal domain is divided in three subdomains (A, B and C). Subdomain A forms a compact bundle, mainly composed by helices, which hangs off the outside of the ring-shaped complex. EM analysis suggests that this domain changes conformation, depending on its phosphorylation state (Chen et al., 2005, Fletcher and Chen, 2006). Subdomain B forms a compact domain, made of three antiparallel β-sheets, which contains a zinc-binding site (Fletcher et al., 2003). Subdomain C contains five antiparallel β-sheets, which form a β-barrel like structure, and represents the core of the N-terminal domain and it is connected to the AAA+ domain through a flexible linker (Fletcher et al., 2003, Liu et al., 2008, Brewster et al., 2008).

The MCM AAA+ domain is constituted of five parallel β-sheets flanked on each side by α-helices, and retains structural features that are common to all AAA+ proteins: a canonical P-loop motif, which contains a conserved Walker A lysine that coordinates and binds the γ phosphate of the ATP, and a Walker B motif, which coordinates Mg^{++} and water to catalyze the hydrolysis of ATP to ADP and phosphate (Brewster et al., 2008, Bae et al., 2009, Erzberger and Berger, 2006).

The C-terminal domain is connected to the AAA+ domain through a flexible linker and adopts a winged helix-turn-helix fold (Barry et al., 2007).

As mentioned before, archaeal MCM proteins have proven to be powerful tools to elucidate essential features of MCM functions and this is linked to their structural organization (Miller and Enemark, 2015).

The eukaryotic MCM2-7 ring might present difficulties in the crystallization process due to multiple orientations that can be adopted in the crystallographic lattice which might look similar at macroscopic level, but not at atomic level because the ring is formed by six different polypeptide chains. On the other hand, archaeal MCM complexes often consist of six identical subunits, meaning all the permutations are identical, at both macroscopic and atomic level, and potentially facilitating crystallization. For this reason, two archaeal MCM proteins, one from the thermophilic archaeon *Sulfolobus solfataricus* (*Sso*MCM) (Liu et al., 2008, Brewster et al., 2008) and another one from *Methanothermobacter thermautotrophicus* (*Mt*MCM) (Fletcher et al., 2003), represent the best studied examples to date. Crystal structures from these two Archaeons provide the vast majority of high-resolution structural information available.

Some archaeal MCM proteins are polymorphic, both hexameric and heptameric architectures have been reported for the same polypeptide, moreover both these assemblies can be found as single rings, double rings or filaments (Figure 4.1B). Electron microscopy reconstruction provide low resolution maps of single hexameric and heptameric full-length MCM rings, as well as double rings and filament from several species of Archaea (Costa et al., 2006a, Costa et al., 2006b, Costa et al., 2008, Gomez-Llorente et al., 2005, Fletcher et al., 2003, Slaymaker et al., 2013). The question on whether all these different assemblies can be present *in vivo* or if they are solely artefacts of overexpression and purification conditions, still remain unsolved and further experiments need to be undertaken to elucidate this topic.

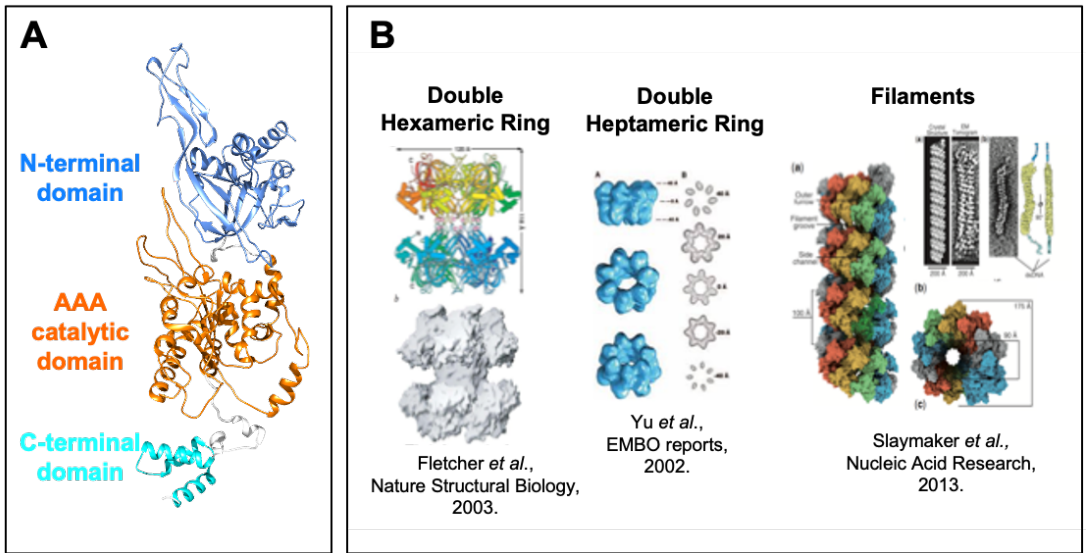


Figure 4.1: Archaeal MCM proteins. A: Homology model for *Pab*MCM obtained using Phyre2; the N-terminal domain is shown in blue, the AAA+ catalytic domain in orange and the C-terminal domain in cyan. Inter-domains linkers are represented in grey. B: Polymorphism of archaeal MCM complex; double hexameric, double heptameric assemblies and filaments were reported by Fletcher *et al.*, Yu *et al.*, and Slaymaker *et al.*, respectively. [Adapted from Fletcher *et al.* (2003), Yu *et al*. (2002), and Slaymaker *et al*. (2013)].

Available evidences suggest that MCMs are commonly associated in a head-to-head double ring configuration and, after origin melting, the double hexamer can separate into two single rings that translocate unwinding the DNA (Remus et al., 2009, Takahashi et al., 2005). However, the mechanism by which MCM loads to origins of replication, melts them and transits from a melted origin to two viable replication forks remains unclear. It is a reasonable assumption that the conformation of MCM has to be dynamic so that the rings can open and reclose to bind DNA, as required at different stages of replication (Costa et al., 2011, Bochman and Schwacha, 2010). It follows that there are two possible scenarios in which MCM associates with genomic DNA. In the first one, MCM is loaded onto dsDNA at the origin of replication, which is subsequently melted (Remus et al., 2009, Takahashi et al., 2005). Whereas, in the second scenario, other replication factors, such as ORC and Cdc6, melt the replication origin and MCM loads afterwards onto ssDNA (Samson et al., 2016, Samson and Bell, 2016, Fernandez-Cid et al., 2013, Frigola et al., 2013, Frigola et al., 2017).

An approach that could be used to dissect the individual steps of MCM mechanism is to use distinct DNA molecules designed to mimic the shape adopted by genomic DNA during individual phases of replication. For instance, in order to examine the wrapping of the origin of replication by head-to-head double ring MCM complexes, we could use a long linear dsDNA as substrate. On the other hand, the configuration adopted by MCM during the process of origin melting could be studied using a “bubble”-shaped DNA and it could be compared to the one assumed during the progression of the replication fork by using a y-shaped DNA substrate which mimics DNA unwinding.

In this chapter, we present the characterization of the archaeal *Pyrococcus abyssi* MCM assembly (*Pab*MCM), in isolation or in complex with distinct DNA substrates, in order to provide a deeper understanding of the mechanism by which archaeal MCMs dynamically loads to origins of replication, melts them and transits from a melted origin to two viable replication forks. Additionally, in this study we want to further investigate the peculiar case of *Pab*MCM, which represents the first evidence that MCM proteins can be found not only as hexamers and heptamers, but also as octameric assemblies (Cannone et al., 2017). Furthermore, in this chapter, *Pab*MCM is used as a tool to provide technical insights on the characterization of heterogeneous and extremely large protein-nucleic acid complexes, by using an integrated approach that combines electron microscopy and small angle scattering studies.

4.2 Materials and Methods

4.2.1 Protein overexpression and purification

Recombinant *Pab*MCM was overexpressed in *E. coli* Rosetta™(DE3) pLysS transformed by heat-shock method. A single colony was picked and used to inoculate a starting culture, which was grown in LB medium containing 50 µg/ml ampicillin and 34 µg/ml chloramphenicol at 37°C o/n. 1 ml of overnight culture was inoculated in 1 L of TB medium supplemented with antibiotics and cells were grown to an OD_{600}=0.6. Protein overexpression was induced by adding 0.2 mM IPTG and culturing the cells at 30°C overnight. Cells were harvested by centrifugation at 5000 rpm on a Beckman Avanti™ J-20 XP centrifuge with JLA 8.1000 rotor for 20 minutes and washed once with PBS; cell pellets were aliquoted and stored at -80°C. Frozen aliquots were thawed and lysed by sonication in 20 ml of lysis buffer (20 mM HEPES pH 7.4, 500 mM NaCl, 400 µl Complete EDTA-free protease inhibitors 50X) followed by 20 min of heat–denaturation at 70°C. Soluble protein extracts were separated from cell pellets by centrifugation at 15000 rpm for 30 min. 6-His tagged *Pab*MCM protein was purified by nickel chromatography using HisTrap FF columns (GE Healthcare) equilibrated in washing buffer (20 mM HEPES pH 7.4, 500 mM NaCl, 20 mM imidazole). Elution was performed with a 20 mM-500 mM imidazole gradient. Fractions containing the protein of interest were tested for DNA contaminations by measuring the $Abs_{260/280}$ ratio. Only fractions with $Abs_{260/280}$ ratio between 0.51 and 0.68 (99% protein) were pulled and concentrated by ultra-filtration using a 15R Vivaspin (30,000 MWCO), to a final concentration of 10 mg/ml prior to size exclusion chromatography on a Superose6™10/300 GL column in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl). Overexpression and purification efficiency were monitored by SDS-PAGE analysis on 10% gels stained with SimplyBlue SafeStain (Invitrogen). Fractions were also analyzed by Western Blot: the membrane was incubated o/n at 4°C with anti-His antibody conjugated with alkaline phosphatase (1:4000) and developed using the SIGMA-FAST™ BCIP/NBT reagent.

4.2.2 Helicase activity assay

Purified *Pab*MCM protein (2 µM) was incubated with annealed y-shaped FAM-DNA (10 nM) at 65°C and 75°C for 30 minutes, 1,4 and 20 hours, in helicase buffer (20 mM HEPES pH 7.4, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA, 3 mM ATP 150 mM NaCl and 100 mM K-glutammate). The sequences of the oligos forming y-shaped FAM-DNA can be found in Table 4.1. Temperature and protein-DNA ratio were chosen according to experimental condition commonly used for helicase activity assays of *Sso*MCM. Adequate DNA concentration was previously tested in order to have sufficient detectable fluorescence. *Pab*MCM helicase activity was tested by running the samples in a 10% native-PAGE and boiled y-shaped FAM-DNA was used as positive control. Imaging was performed using a Typhoon FLA 9500 phospho-imager.

OLIGO name	Sequence
OLIGO_1(FAM5')	5'-CAAGCAGTCCTAACTAGGCAGAGTCCCCCACCTAACTTTAA -3'
OLIGO_2	5'-TTAAAGTTAGGTGGGGGACTCTGCCTCAAGACCACTGACAACGTGACCGCAGCCCACCTG-3'

Table 4.1: Oligos used to form y-shaped FAM-DNA. The sequence of the fluorescent tag (FAM) at the 5' end of OLIGO_1 is underlined. OLIGO_2 is partially complementary to OLIGO_1, forming a y-shaped DNA that recreates the shape of the replication fork.

4.2.3 Protein-DNA complex assembly

DNA oligos (Table 4.2) were annealed to form different shapes of DNA substrates, dsDNA (OLIGO_A + OLIGO_B) and y-shaped DNA (OLIGO_A + OLIGO_C). 50 µl of each oligo (100 µM), forming the DNA substrate, were added into a PCR tube and mixed before being boiled at 95°C for 10 minutes, annealing was achieved o/n by incubation of the PCR tubes into boiling water, allowing a slow and constant decrease in temperature. 500 µl of nickel-purified *Pab*MCM (10 mg/ml) were incubated with 100 µl of different DNA substrates (50 µM) for 1 hour at room temperature prior to size exclusion chromatography on a Superose6™10/300 GL column in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl). Purification was monitored by SDS-PAGE analysis on 10% gels stained

with SimplyBlue SafeStain (Invitrogen) and changes into *Pab*MCM assemblies were monitored by BN-PAGE using 4-12% NativePAGE Bis-Tris protein gels (Invitrogen).

OLIGO name	Sequence
OLIGO_A	5'-TATGGCTGTCACTGTAATCCCGGTATATTTAGCTCATGTACATGAGTCTATAAATAAC-3'
OLIGO_B	5'-GTTATTTATAGACTCATGTACATGAGCTAAATATACCGGGATTACAGTGACAGCCATA-3'
OLIGO_C	5'-TAGCTCATGTACATGAGTCTATAAATAACAATATACCGGGATTACAGTGACAGCCATA-3'

Table 4.2: Oligos used to form DNA substrates. OLIGO_A and OLIGO_B were annealed to form dsDNA, while OLIGO_A and OLIGO_C were annealed to form y-shaped DNA.

4.2.4 Small angle X-ray scattering (SAXS) and modelling

SAXS data for *Pab*MCM in isolation and in complex with distinct DNA substrates (in absence or presence of ATPγS) was collected at B21, Diamond Light Source (Harwell, UK). 55 µl of each protein sample (~10 mg/ml) was loaded onto a Superose6 column (GE Healthcare), controlled by an Agilent HPLC system, coupled to an in-vacuum SAXS flow cell. HPLC-SAXS traces were processed using ScÅtter. Data was analysed using different strategies. *Ab initio* bead models were obtained using ScÅtter and DAMMIN. Ensemble models of single hexameric rings in open conformation, closed single heptameric rings and single octameric rings were generated using EOM: a pool of 10,000 asymmetric independent models was generated, based on the sequence of *Pab*MCM and on constraints we generated by homology modeling using Phyre2. The constraints used as input for EOM correspond to the three main *Pab*MCM domains (N-terminal, AAA+ and C-terminal). Particular care was taken in order to build the ring assembly of N-terminal domains formed by six, seven or eight subunits. Three independent N-terminal rings models were built using Chimera by docking the N-terminal domains onto the EM density maps of *Pab*MCM and onto the crystallographic structure of

*Sso*MCM. Theoretical SAXS curves of individual EOM models were calculated using CRYSOL. Figures were prepared using ScÅtter and SasView.

4.2.5 200 kV cryo-electron microscopy (cryo-EM)

Cryo-EM and single particle analysis (SPA) were coupled to study the structure of *Pab*MCM. The cryo-EM data collection was performed on Quantifoil 1.2-1.3 copper grids covered with a 2nm layer of carbon, plunge-frozen into liquid ethane using a Vitrobot instrument. The data collection was carried under low-dose mode at nominal magnification of 50,000x, at a final sampling of 1.51 Å/pixel at the specimen level, using a FEI F20 FEG microscope, equipped with a TemCam-F816 (8kx8k) CCD camera. Data were processed using RELION and EMAN2 softwares. From the micrographs dataset, ~34,000 single particles (SP) were manually picked using 250 Å masks; the single particle images were extracted into boxes 200x200 pixels, re-scaled at 96x96 pixels and band-pass filtered. The dataset was subjected to sequential rounds of alignments and classification in order to improve the resulting class-average images. The original dataset was divided into sub-datasets on the basis of the class-averages obtained with the 2D classification. Firstly, two sub-datasets were created for *Pab*MCM in single ring conformation (~26,000 SP) and in double ring conformation (~8,000 SP). Secondly, the sub-dataset of the single ring *Pab*MCM was split into two sub-datasets grouping heptametric or octameric (~16,000 SP) single particles. Only the single particle images from the octameric single ring dataset were used for the following 3D reconstruction processes. Figures were prepared with UCSF Chimera and fitting copies of the crystallographic structure of *Sso*MCM of (PDB ID: 3F9V) into the cryo-EM density maps obtained.

4.2.6 300 kV cryo-electron microscopy (cryo-EM)

Cryo-EM and single particle analysis (SPA) were coupled to study the structure of *Pab*MCM in complex with y-shaped DNA and 5 mM ATPγS. The cryo-EM data

collection was performed on Quantifoil 1.2-1.3 copper grids, plunge-frozen into liquid ethane using a Vitrobot instrument. The data collection was carried under low-dose mode at nominal magnification of 50,000x, at a final sampling of 1.06 Å/pixel at the specimen level, using a FEI Titan Krios microscope, equipped with a Gatan Quantum K3 Summit image filter/detector. Data was processed using RELION and EMAN2 softwares.

4.3 Results and Discussion

4.3.1 *Pab*MCM has no significant helicase activity *in vitro*

E. coli Rosetta™(DE3) pLysS cells were transformed by heat-shock method with pTWO-E vector containing *Pab*MCM insert. Protein overexpression was induced by addition of 0.2 mM IPTG and carried out o/n at 30°C. Cells were then lysed by sonication and the lysate was incubated at 70°C for 20 minutes.

Heat-denaturation represents the first step in the purification protocol of *Pab*MCM as this archaeal protein is highly thermo-stable and does not aggregate, whereas the majority of *E. coli* proteins become insoluble after heat treatment and the aggregates can be removed by a simple centrifugation step. *Pab*MCM protein was then purified by nickel chromatography in high-salt buffer in order to remove nucleic acid contamination. Elution was performed with an imidazole gradient, the fractions containing protein were analysed by 10% SDS-PAGE analysis and then pooled and concentrated to 10 mg/ml prior to size exclusion chromatography (SEC) using a Superose6 column.

From the SEC trace in Figure 4.2A it is possible to notice that the main peak, eluting after ~13.5 ml, is characterized by a shoulder on its left-hand side, indicating that a bigger assembly elutes after ~11.5 ml and the two peaks are not resolved. The apparent molecular weight of the main peak of the SEC profile of *Pab*MCM is 575 kDa, suggesting that this protein is likely to form an octameric assembly. Moreover, the apparent molecular weight of the left-hand side shoulder peak (~ 1 MDa) is consistent with the one of a double ring assembly. From the gel in Figure 4.2B it is possible to observe that the main band corresponds to a molecular weight consistent to the one of the *Pab*MCM monomer (74 kDa). However, a small amount of degradation products could be detected when the gel was overloaded. Removal of these contaminants was carried out using 100 K MWCO spin-concentrators before further experiments.

The helicase activity of *Pab*MCM was tested by incubating the purified protein with fluorescently labelled y-shaped DNA that mimics the replication fork, in the presence of ATP, and by native-PAGE analysis. More specifically, using this

assay, if the protein has helicase activity, we would observe a band-shift of the fluorescent DNA substrate at lower molecular weight compared to the one of free duplexed y-shaped DNA. Temperature and protein-DNA ratio, as well as buffer composition, were chosen accordingly to experimental condition commonly used for helicase activity assays of *Sso*MCM. Adequate DNA concentration was previously tested in order to have sufficient detectable fluorescence. As shown in Figure 4.3, *Pab*MCM has very low helicase activity *in vitro*, detectable only after 20 hours of incubation at 70°C. Moreover, it is possible to observe fluorescence at the bottom of the wells were protein and DNA are simultaneously present, indicating high-molecular weight complex assembly. *P. abyssi* is a hyper-thermophilic archaeon and its optimal temperature habitat is 95°C. It follows that, since DNA is denaturated *in vitro* at 95°C, lower temperatures and longer incubation times had to be used in order to test *Pab*MCM helicase activity *in vitro*. Consequently, the extremely low helicase activity detected for *Pab*MCM protein could also be associated to limitations in the choice of *in vitro* experimental conditions.

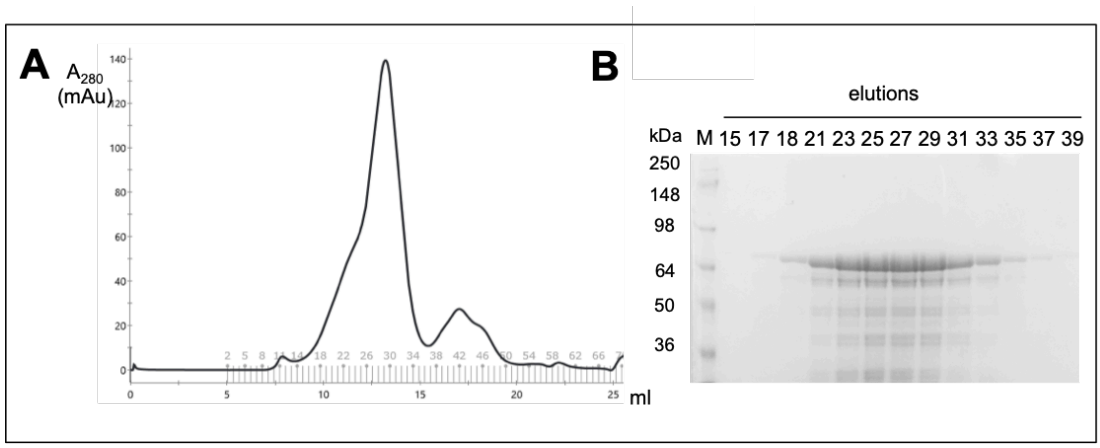


Figure 4.2: SEC purification step of *Pab*MCM. A: Size exclusion chromatography trace of *Pab*MCM in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl) using a Superose6™10/300 GL column at 0.4 ml/min flow-rate. B: SimplyBlue SafeStain staining of 10% SDS-PAGE analysis of the elutions of SEC purification step of *Pab*MCM.

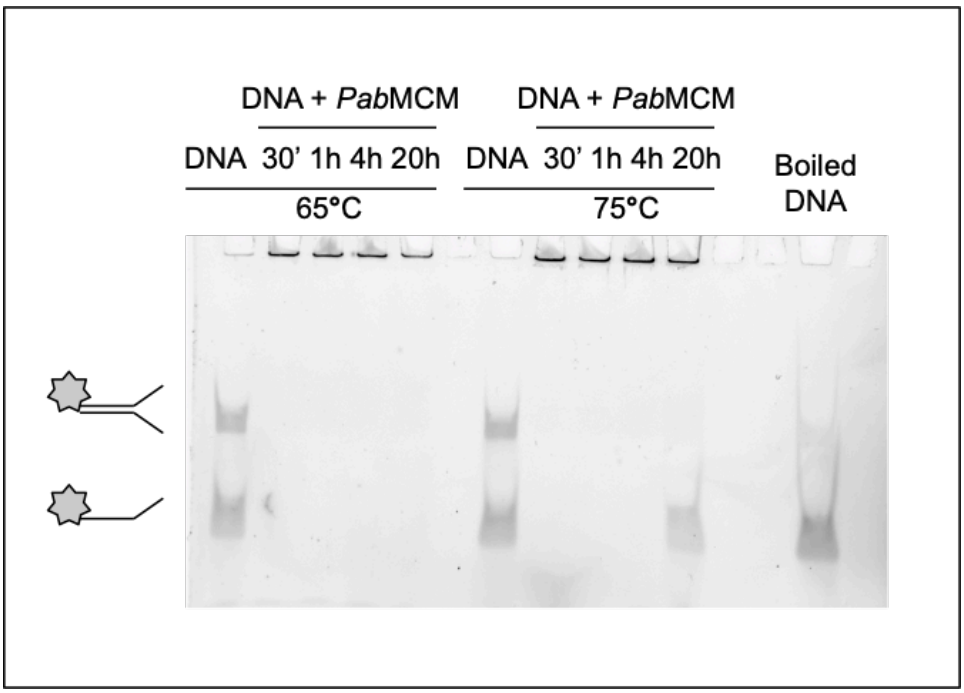


Figure 4.3: Helicase activity assay of *Pab*MCM protein. Purified *Pab*MCM protein (2 µM) was incubated with y-shaped FAM-DNA (10 nM) at 65°C and 75°C for 30 minutes, 1,4 and 20 hours, in helicase buffer (20 mM HEPES pH 7.4, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA, 3 mM ATP 150 mM NaCl and 100 mM K-glutammate) and loaded into a 10% native-PAGE. Imaging was performed using a Typhoon FLA 9500 phospho-imager.

4.3.2 *Pab*MCM is a heterogeneous complex

Most archaeal MCM proteins are polymorphic and both hexameric and heptameric architectures have been reported for the same polypeptide. Moreover, both these assemblies can be found as single rings, double rings or filaments. *Pab*MCM represents the first evidence that MCM proteins can be found not only as hexamers and heptamers, but also as octameric assemblies (Cannone et al., 2017).

As shown in Figure 4.4, *Pab*MCM forms numerous assemblies. There are three main bands between 480 kDa and 720 kDa, which are likely to correspond to single ring assemblies in hexameric, heptameric and octameric configuration. However, other weaker bands can be detected at molecular weights consistent with the ones of single rings and these could correspond to different conformations of this dynamic complex.

On the other hand, a single band is present between 720 kDa and 1048 kDa, which is likely to correspond to a double ring complex. The molecular weight would suggest a double hexameric assembly (888 kDa). However, the presence or absence of double heptamers and double octamers would be difficult to detect using this gel electrophoresis. It follows that, using this analysis, we cannot exclude the possibility that also *Pab*MCM double ring is an heterogeneous complex, as demonstrated for the single ring assembly.

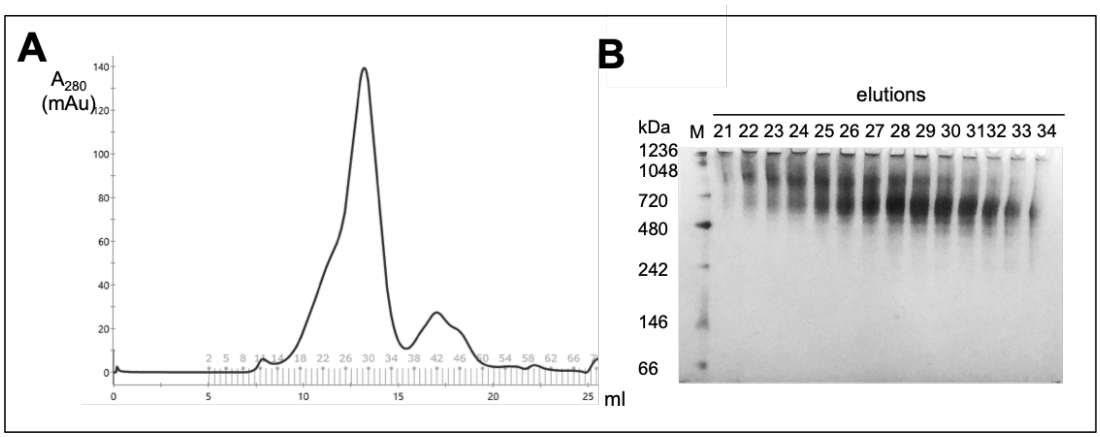


Figure 4.4: Heterogeneity analysis of *Pab*MCM. A: Size exclusion chromatography trace of *Pab*MCM in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl) using a Superose6™10/300 GL column at 0.4 ml/min flow-rate. B: BN-PAGE analysis of the elutions from SEC purification step of *Pab*MCM in a 4-12% Bis-Tris NativePAGE, gel was run for 3 hours at 150 V and de-stained o/n in 30% ethanol and 10% acetic acid.

Cryo-EM and single particle analysis (SPA) were coupled to study the structure of *Pab*MCM and its heterogeneity (Figure 4.5). From the CTF corrected micrographs dataset, ~34,000 single particles (SP) were manually picked using 250 Å masks; the single particle images were extracted into boxes 200x200 pixels, re-scaled at 96x96 pixels and band-pass filtered. The dataset was subjected to sequential rounds of alignments and 2D classification in order to improve the resulting class-average images. The original dataset was divided into sub-datasets on the basis of the class-averages obtained with the 2D classification. Firstly, two sub-datasets were created for *Pab*MCM in single ring conformations (~26,000 SP) and in double ring conformation (~8,000 SP). Secondly, the sub-dataset of the single ring *Pab*MCM was split into two sub-datasets grouping heptametric (~10,000 SP) or octameric (~16,000 SP) single particles. From the 2D class-averages obtained it is not clear whether or not single ring hexameric assemblies are present in our dataset. In fact, some of the class-averages belonging to the heptameric ring sub-dataset might instead represent a hexameric ring in an open configuration. This derives from the fact that some densities are in different axial planes and they could be either single subunits or different domains of the same *Pab*MCM monomer. For this reason, only the single particle images from the octameric single ring dataset and for the double ring were used for the following 3D reconstruction. Initial models were generated with EMAN2 without imposing symmetry and checked by inspecting the 3D model and the correspondence between the class averages and the re-projections of the initial models. The choice of avoiding to impose symmetries comes from the fact that previous EM structural studies of other MCM proteins from different organisms show that these proteins assemble as skewed rings (Lyubimov et al., 2012). The models for the single octameric ring and for the double ring, obtained after 3D classification and refinement in RELION, were manually fitted with copies of the 4.35 Å resolution crystallographic structure of a near full-length *Sso*MCM (PDB ID: 3F9V), whose AAA+ domain presents high sequence identity with the one of *Pab*MCM. On the other hand, the N-terminal domain of *Pab*MCM reveals low sequence similarity with previously reported *S. solfataricus* and *M. Thermoautrophicus* MCM proteins (Cannone et al., 2017), even though the secondary structure of the N-terminal domain is highly conserved amongst Archaea and Eukaryotes. The density map obtained for the single octameric ring shows that indeed eight copies of 3F9V can

be fitted into the EM map (Figure 4.6). Moreover, two subunits are in a different axial plane compared to the rest of the ring, which is indeed in a skewed configuration. The density map obtained for the double ring shows that sixteen copies of 3F9V could be fitted in the map, suggesting a head-to-head assembly of octameric rings (Figure 4.7).

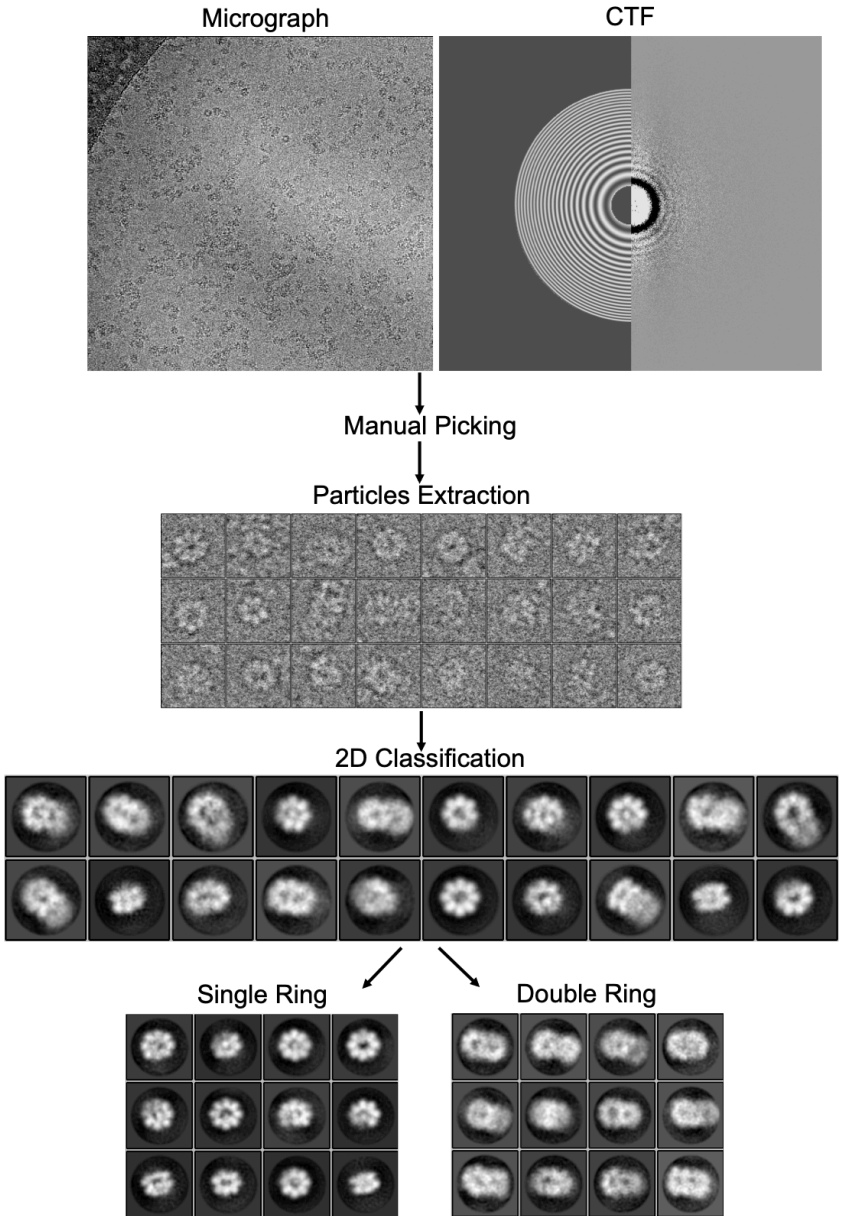


Figure 4.5: Cryo-EM coupled to single particle analysis of *Pab*MCM. Micrographs were collected at nominal magnification of 50,000x at 1.51 Å/pixel, using a FEI F20 FEG microscope equipped with a TemCam-F816 (8kx8k) CCD camera. Micrographs were CTF corrected and ~34,000 single particles were manually picked using 250 Å masks and subjected to several round of classification to then be divided into sub-datasets corresponding to *Pab*MCM in single ring conformation (~26,000 SP) and in double ring conformation (~8,000 SP).

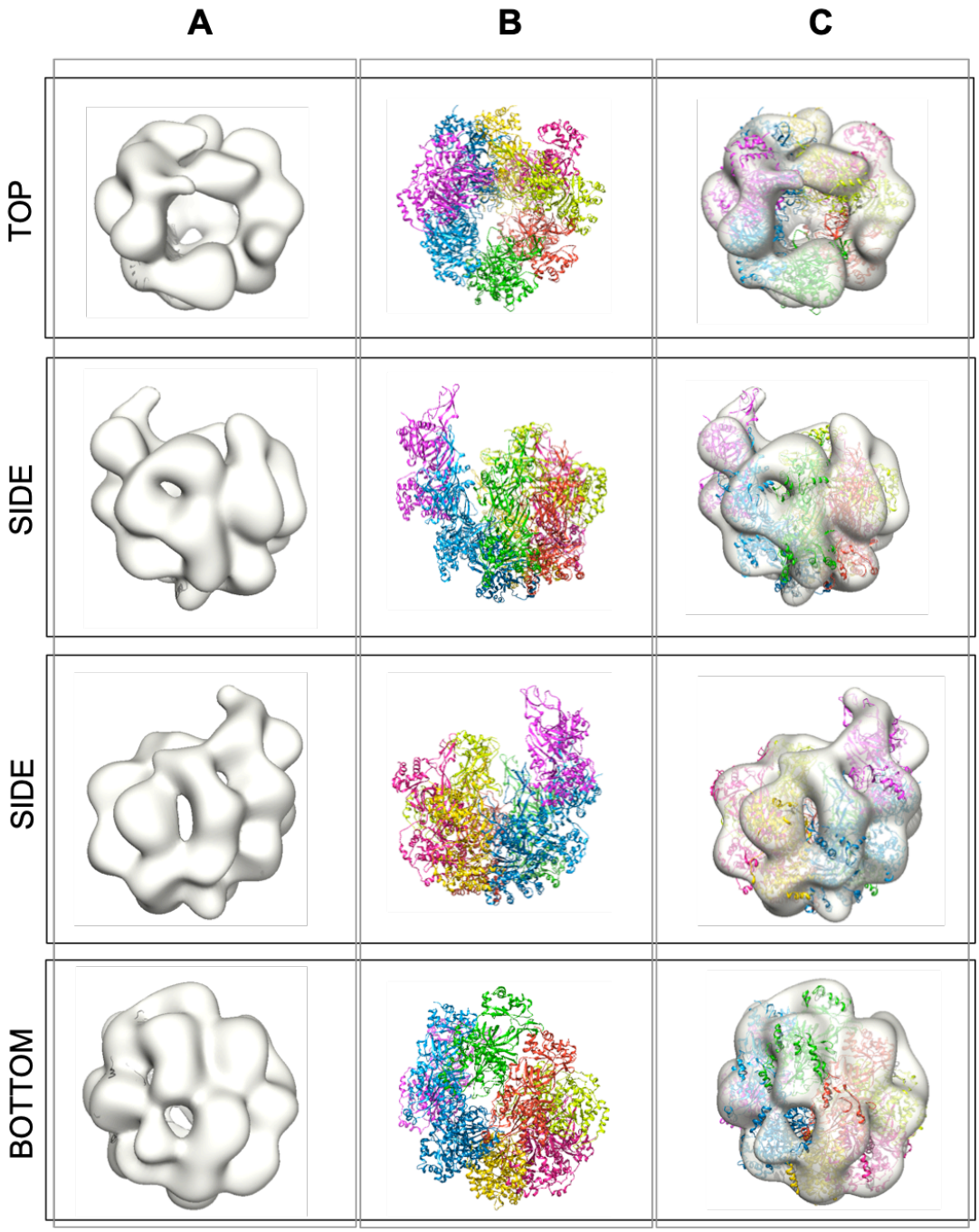


Figure 4.6: Cryo-EM density map of *Pab*MCM single octameric ring. A: ~16,000 single particles were used to generate a 3D model of *Pab*MCM in single octameric ring assembly. B: 8 truncated *Sso*MCM monomers (PDB ID: 3F9V) were fitted into the cryo-EM density map. C: Docking of truncated *Sso*MCM monomers into the 3D model of *Pab*MCM single octameric ring assembly.

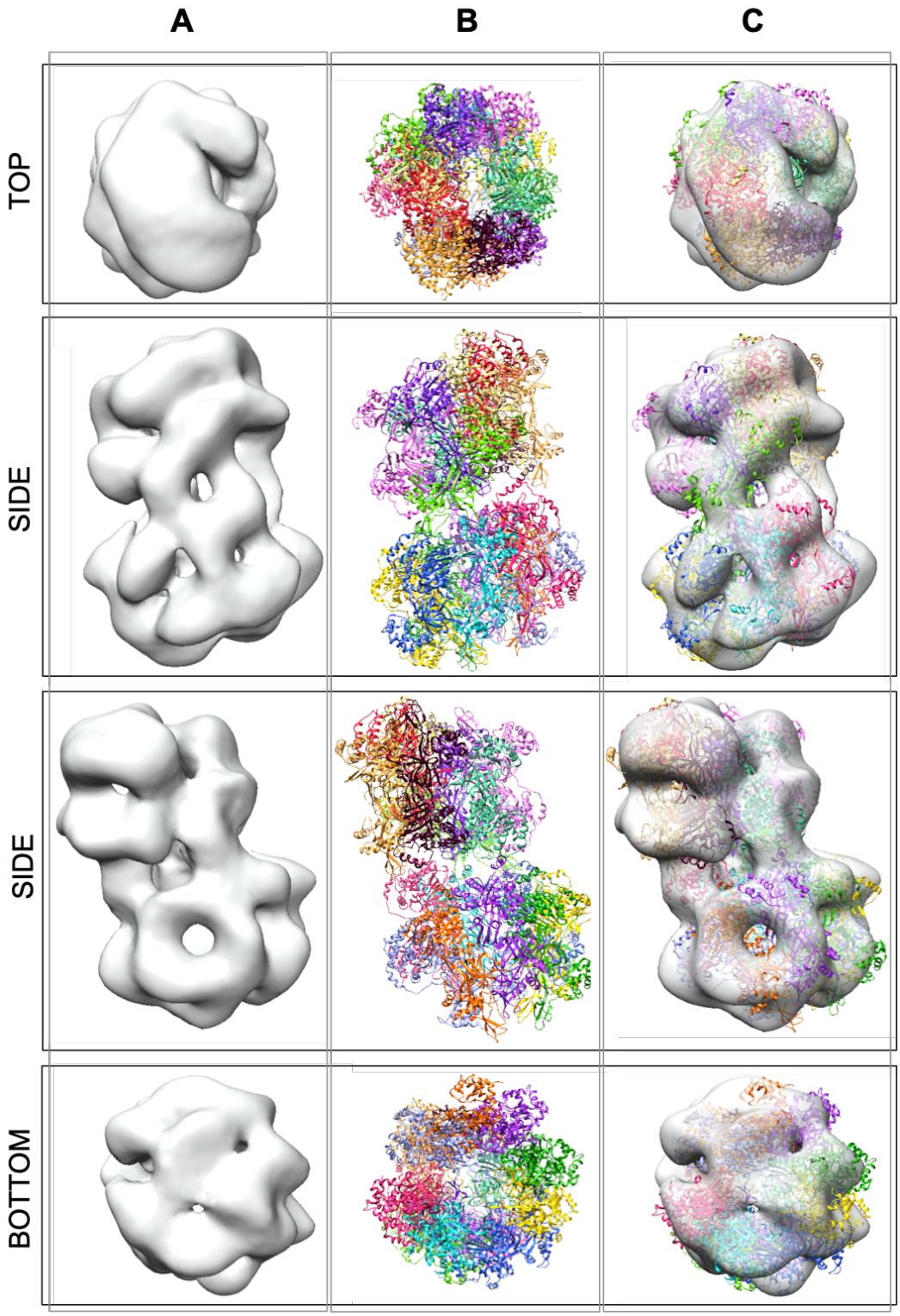


Figure 4.7: Cryo-EM density map of *Pab*MCM double octameric ring. A: ~8,000 single particles were used to generate a 3D model of *Pab*MCM in double octameric ring assembly. B: 16 truncated *Sso*MCM monomers (PDB ID: 3F9V) were fitted into the cryo-EM density map. C: Docking of truncated *Sso*MCM monomers into the 3D model of *Pab*MCM double octameric ring assembly.

4.3.3 *Pab*MCM forms dynamic rings

The first evidence that *Pab*MCM can adopt several configurations in solution is given by the presence of multiple bands consistent with molecular weights of single ring assemblies. Moreover, cryo-EM coupled to single particle analysis shows that *Pab*MCM can assemble as single octamers, heptamers and probably hexameric rings in open conformation. Small angle X-ray scattering was used in order to provide more information about the behaviour of the complex in solution. SAXS data were collected at B21, Diamond Light Source (Harwell, UK). The protein sample was loaded onto a Superose6 column (GE Healthcare), controlled by an Agilent HPLC system coupled to an in-vacuum SAXS flow cell. HPLC-SAXS traces were processed using ScÅtter. Particular care was taken when analysing the HPLC-SAXS traces, in order to avoid using frames with non-ideal separation of single and double ring assemblies. The low number of frames used, and therefore the low concentration, is the basis of the high noise of the data, especially at high *q*.

The buffer subtracted experimental SAXS intensity curve of *Pab*MCM in single ring assembly is shown in purple in Figure 4.8A and the correspondent *ab initio* bead model, represented in purple, was obtained using ScÅtter and DAMMIN. The R_g calculated from the Guinier analysis is 68.99 Å, whereas the one obtained in real space from the *p(r)* distribution is 70.01 Å. The D_{max} is 283 Å and the χ^2 of the fitting is 0.37. Even though the DAMMIN model fits well to the SAXS data (χ^2 = 0.71), it is possible to notice the presence of extra density compared to the bead model generated from the EM density map of the single ring in octameric assembly, represented in grey. Moreover, the theoretical SAXS curve calculated for the EM model, shown in black in the graph, presents more enhanced features compared to the experimental data. The same approach was used to study *Pab*MCM in double ring assembly and its experimental SAXS intensity curve is shown in blue in Figure 4.8B. The R_g calculated from the Guinier analysis is 110.46 Å, whereas the one obtained in real space from the *p(r)* distribution is 113.06 Å. The D_{max} is 418 Å and the χ^2 of the fitting is 0.53. The DAMMIN model, shown in blue, fits well the SAXS data (χ^2 = 0.73), but also in this case it is possible to observe the presence of an extra density.

Additionally, the theoretical SAXS curve calculated for the EM model of *Pab*MCM double ring, shown in black in the graph, also presents more distinct features compared to the experimental curve.

The extra densities observed in the SAXS beads models for *Pab*MCM single ring and double ring are not likely to derive from aggregation or inter-particle interference as fractions from SEC purification were previously assessed with DLS measurements, in order to identify regions of the peaks containing only one species. Consequently, we hypothesized that the dissimilarity between the solution scattering models and the cryo-EM models could be explained by the intrinsic difference in data processing of these two techniques. In fact, while 2D class averaging and further subsets selection of different conformations is possible in cryo-EM analysis, the SAXS profile is the result of the scattering pattern of all molecules in solution and the identification of different conformations in solution can be particularly challenging. Importantly, SAXS shows that both these systems are characterized by a certain degree of flexibility, as shown by their Kratky plot in Figure 4.8C, and this could be the reason why the DAMMIN models exhibit extra densities.

In order to gather more information about these systems, the three-dimensional models of *Pab*MCM protein in single ring assembly were obtained by combining homology modelling analysis and small angle X-ray scattering data.

An Ensemble Optimization Method (EOM) approach was used in order to generate SAXS models for single rings in heptameric, octameric and open hexameric configuration. A pool of 10,000 asymmetric independent models was generated for each type of complex, based on the sequence of *Pab*MCM and on constraints we generated by homology modeling using Phyre2. The constraints used as input for EOM correspond to the three main *Pab*MCM domains (N-terminal, AAA+ and C-terminal). Particular care was taken in order to build the ring assemblies of N-terminal domains formed by six, seven or eight subunits. This was achieved by using Chimera and by docking the N-terminal domains of *Pab*MCM into the EM density maps of *Pab*MCM and into the crystallographic structure of *Sso*MCM. In the EOM analysis, the N-terminal ring was fixed in the configuration obtained from the docking, while positions of AAA+ and C-terminal domains were allowed to be freely modelled by EOM, without imposing any symmetry. Figure 4.9A shows the three EOM models obtained for the octameric

single ring assembly and their calculated SAXS curves, as well as their comparison with the experimental SAXS curve (in purple). Similarly, Figures 4.9B and 4.9C show the EOM models obtained for the heptameric and for the hexameric single ring assemblies, respectively.

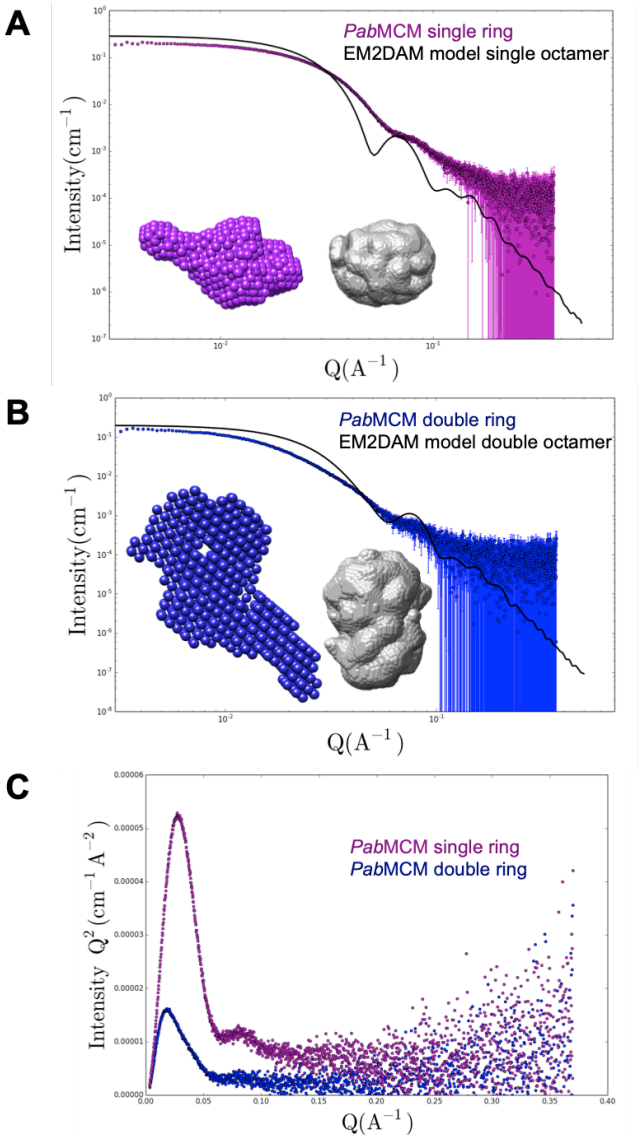


Figure 4.8: Comparison of SAXS and EM characterization of *Pab*MCM in single and double ring assemblies. A: Experimental SAXS intensity curve of *Pab*MCM in single ring assembly (purple) and theoretical SAXS curve calculated from the cryo-EM model of *Pab*MCM in single ring octameric assembly (black), correspondent beads models are shown in purple and grey. B: Experimental SAXS intensity curve of *Pab*MCM in double ring assembly (blue) and theoretical SAXS curve calculated from the cryo-EM model of *Pab*MCM in double ring octameric assembly (black), correspondent beads models are shown in blue and grey. C: Kratky plot analysis of *Pab*MCM in single (purple) and double (blue) ring assemblies. Figures were prepared using SasView and ScÅtter, beads size of the models increased post-analysis for clarity in figure.

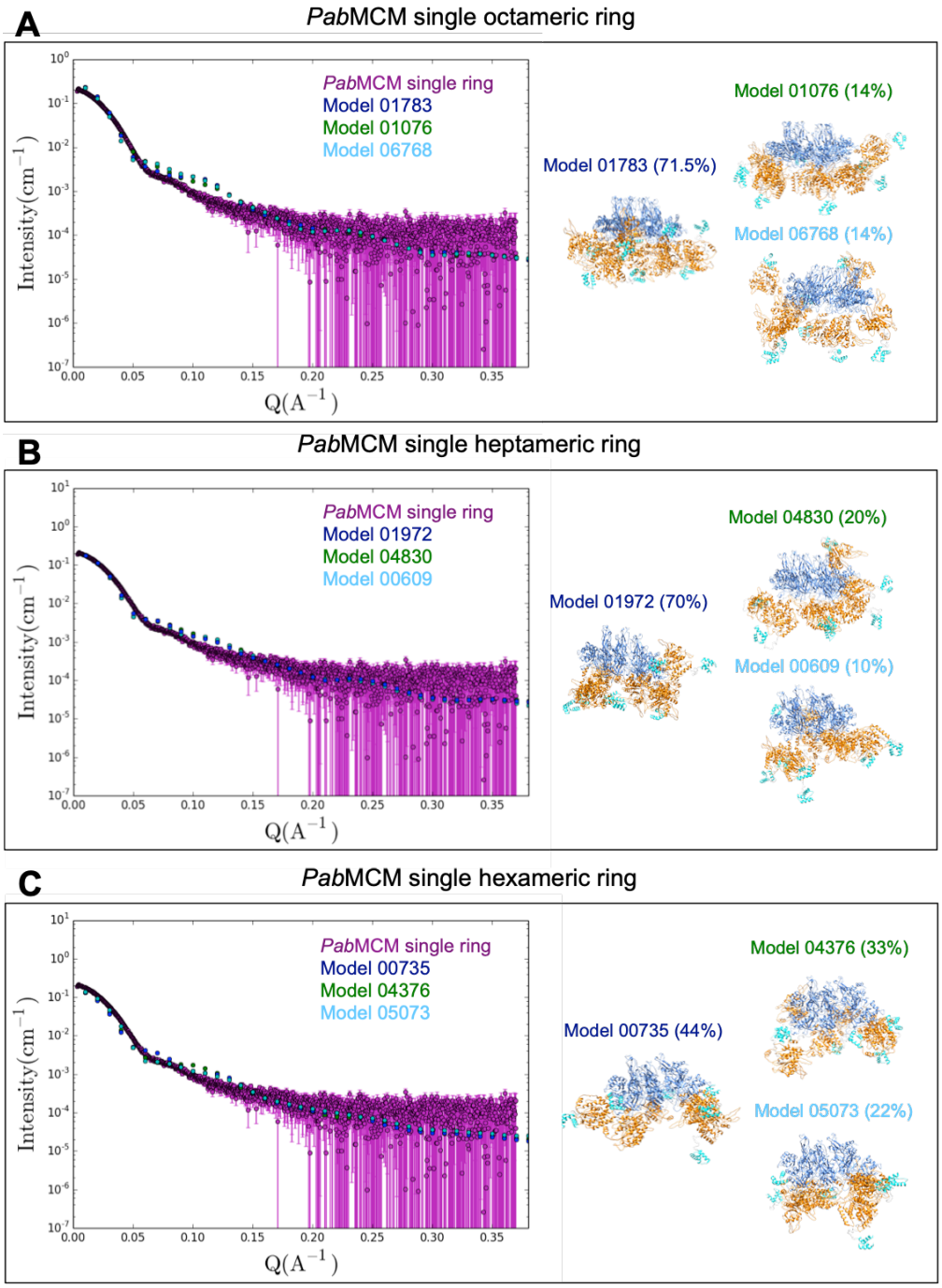


Figure 4.9: EOM modelling of *Pab*MCM in single ring assemblies. A: Experimental SAXS intensity curve of *Pab*MCM in single ring assembly (purple) and theoretical SAXS curve calculated from the EOM models generated for *Pab*MCM in single ring octameric assembly (blue, green, cyan), correspondent EOM models are shown in the figure. B: Experimental SAXS intensity curve of *Pab*MCM in single ring assembly (purple) and theoretical SAXS curve calculated from the EOM models generated for *Pab*MCM in single ring heptameric assembly (blue, green, cyan), correspondent EOM models are shown in the figure. C: Experimental SAXS intensity curve of *Pab*MCM in single ring assembly (purple) and theoretical SAXS curve calculated from the EOM models generated for *Pab*MCM in single ring hexameric assembly in open conformation (blue, green, cyan), correspondent EOM models are shown in the figure. Figures were prepared using SasView.

As shown in Figure 4.9, EOM generated three models for each single ring assembly and, for each of them, the model with the highest percentage in the scattering contribution was chosen to perform further analysis.

The calculated scattering curves of the single rings in different configurations were compared to the experimental SAXS intensity curve (Figure 4.10). As previously described, BN-PAGE and cryo-EM analysis indicate that *Pab*MCM is a heterogeneous complex and can assemble as single hexamers, heptamers and octamers. Moreover, our EOM models fit quite well with our experimental data ($X^2_{hexamer}$ = 4.768, $X^2_{heptamer}$= 3.888, $X^2_{octamer}$ = 10.87).

Therefore, we decided to use OLIGOMER in order to obtain a quantification of the scattering contribution of each model to the total scattering intensity. The software was fed with the nine EOM models obtained for *Pab*MCM single ring (three for each type of complex) and the scattering data. Surprisingly, the output given by this software suggests that only the three EOM models for the single octameric ring contribute to the total scattering intensity, as simplified in Figure 9. Conversely, the calculated X^2 of the three ensembles of EOM models calculated for *Pab*MCM would suggest that the octameric assembly is the less reliable. Another discrepancy between the output from EOM and from OLIGOMER analysis consists in the different percentages of scattering contribution from the three models for the single ring in octameric assembly. Specifically, according to EOM analysis, the model that has the biggest impact on the scattering intensity is “model 01783” (71.5 %), whereas the other two models contribution is 14 % only. On the other hand, OLIGOMER analysis suggests that the “model 01076” is the major species in solution (1 ± 0.695), while the fraction of “model 01783” is only 0 ± 0.664 and the one for “model 06768” is 0 ± 0.693.

Moreover, the R_g calculated with these softwares are inexplicably slightly different and this could be at the basis of the incongruity of the analysis.

For what concerns the *Pab*MCM double ring complex, an EOM approach could not be adopted given the complexity of the system. Instead, a simplified analysis has been performed by fitting two identical copies of each EOM model obtained for the octameric assembly in a head-to-head configuration into the EM density map of the double ring.

SAXS curves were calculated from the models obtained and compared to the experimental data for *Pab*MCM in double ring assembly (Figure 4.11). This

evaluation shows a good fitting with the experimental data. However, further modelling, where the two rings of a double ring assembly are in different conformations or in heptameric and hexameric assembly, was not carried out given the high number of possible combinations and the limited information gain.

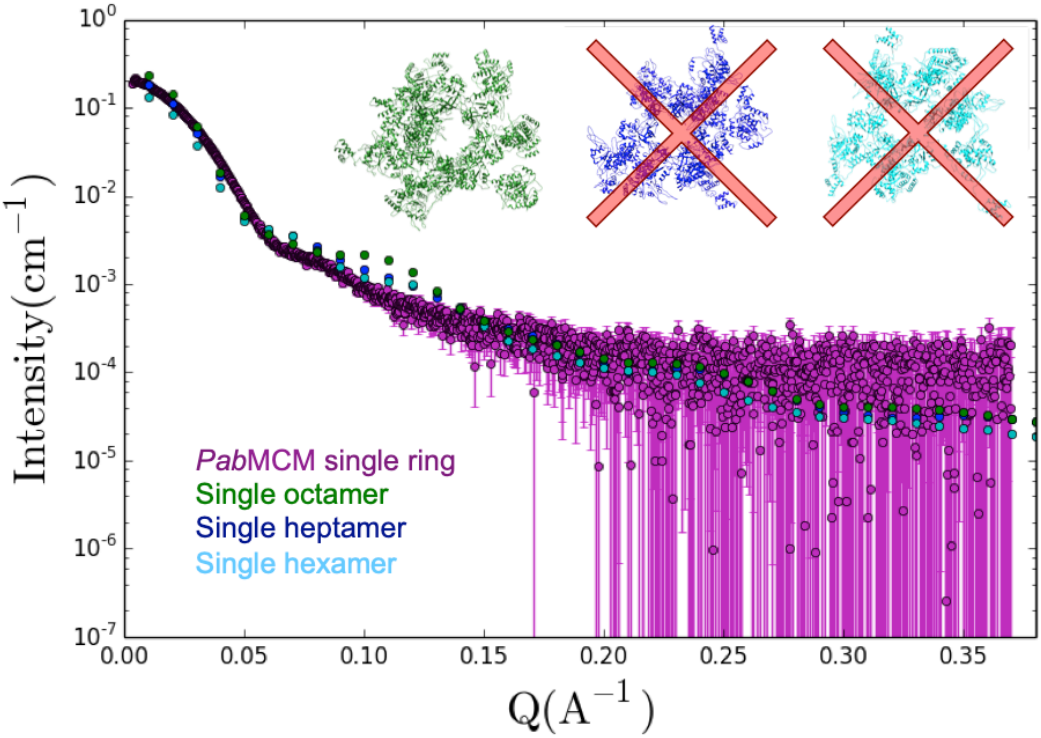


Figure 4.10: Analysis of *Pab*MCM single ring heterogeneity by SAXS. Comparison of the experimental SAXS intensity curve of *Pab*MCM in single ring assembly (purple) and theoretical SAXS curve calculated from the most abundant EOM models generated for *Pab*MCM in single ring octameric (blue), heptameric (green) and hexameric (cyan) assemblies. Correspondent EOM models are shown in the figure. Red crosses highlight the conformations excluded from OLIGOMER analysis. Figures were prepared using SasView.

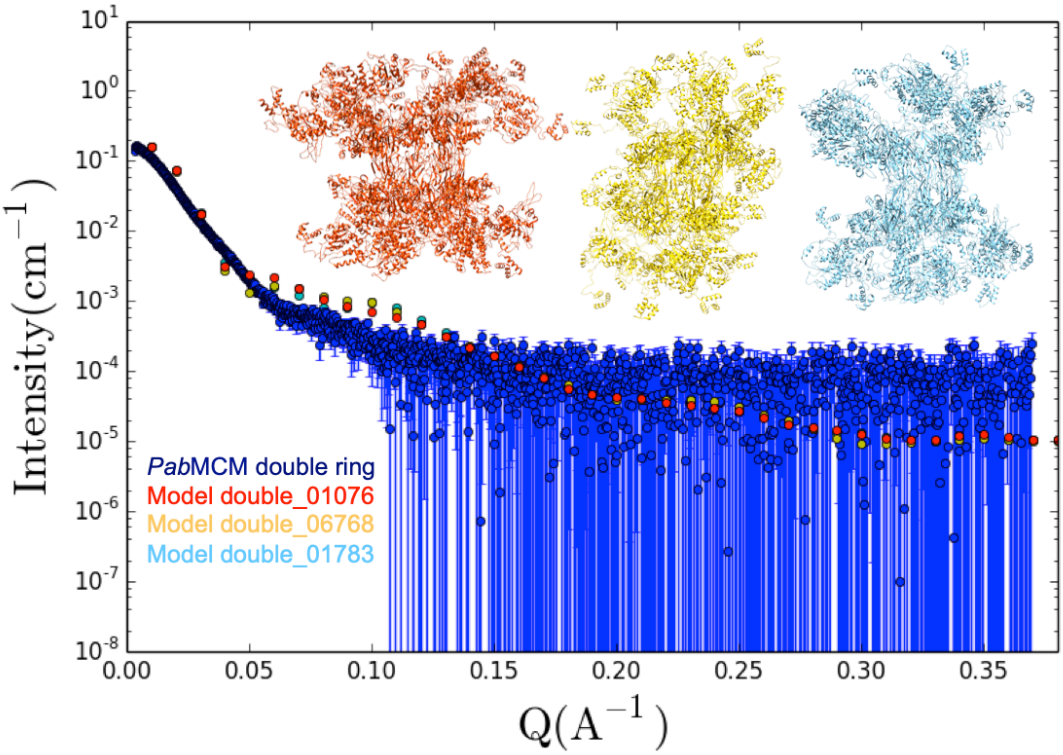


Figure 4.11: SAXS analysis of *Pab*MCM octameric double ring assembly. Comparison of the experimental SAXS intensity curve of *Pab*MCM in double ring assembly (blue) with the theoretical SAXS curves calculated from double ring assemblies generated using UCSF Chimera with *Pab*MCM single ring octameric EOM models as input. Correspondent models and curves are shown in red, yellow and light blue. Figures were prepared using SasView.

4.3.4 DNA binding of y-shaped DNA induces conformational changes on *Pab*MCM complex

In order to dissect individual steps of MCM mechanism and provide deeper understanding of its role in DNA replication, we used distinct DNA molecules designed to mimic the shape adopted by genomic DNA during the different phases of replication. A long linear dsDNA substrate was used to examine whether a head-to-head double ring MCM complex performs the wrapping of the origin of replication. On the other hand, the configuration adopted by MCM during the process of origin melting could be studied using a "bubble"-shaped DNA and it could be compared to the one assumed during the progression of the replication fork by using a y-shaped DNA substrate which mimic the process of DNA unwinding. Preliminary experiments were performed using "bubble"-shaped DNA; however, the low-resolution information obtained with SAXS measurements failed in the detection of MCM rearrangements. For this reason, data for the complex between *Pab*MCM and a "bubble"-shaped DNA substrate are not shown in this thesis and only results for a complex with dsDNA and y-shaped DNA are shown.
Nickel-purified *Pab*MCM was incubated with the different DNA substrates prior to size exclusion chromatography on a Superose6™10/300 GL column. Complex formation was monitored by $A_{260/280}$ measurements and BN-PAGE analysis in a 4-12% Bis-Tris NativePAGE gel.
The comparison between the SEC traces of *Pab*MCM in absence or presence of y-shaped DNA is shown in Figures 4.12A and 4.12C, respectively. As it can be seen from the chromatogram, the binding of the DNA substrate induces a change in the shape of the peak, as the left-hand side shoulder peak, likely to correspond to *Pab*MCM in double ring assembly, seems to increase and to give rise to a non-resolved double peak. However, when comparing the BN-PAGE analysis for *Pab*MCM in absence or presence of y-shaped DNA (Figures 4.12B and 4.12D), it is possible to notice that the fractions corresponding to *Pab*MCM in single ring assembly are shifted towards higher molecular weights when the DNA substrate is added to the mix, suggesting that *Pab*MCM single ring binds y-shaped DNA. Moreover, a reduction of the number of fractions containing *Pab*MCM in double ring assembly can be detected, suggesting that a portion of double rings dissociates in order to elicit DNA binding. On the other hand, *Pab*MCM in double

ring assembly is still present and the distribution of the fractions containing this species suggests that it might become more compact, probably because of DNA binding. Interestingly, upon addition of DNA substrate, it is possible to observe the appearance of a new band, at ~300 kDa, in the elution fractions corresponding to the right-hand side part of the major peak, probably due to partial assembly and/or disassembly of the complex.

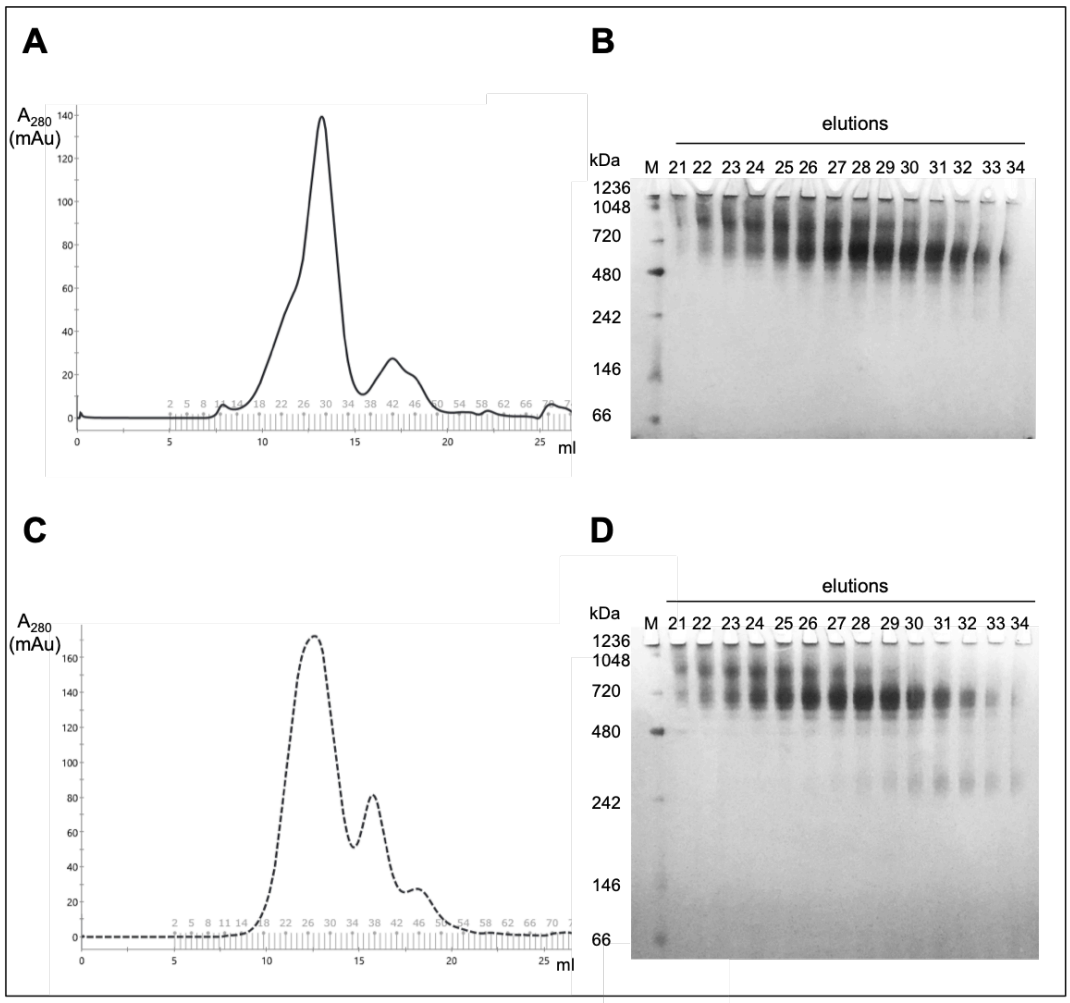


Figure 4.12: Heterogeneity analysis of *Pab*MCM in absence or presence of y-shaped DNA. A: Size exclusion chromatography trace of *Pab*MCM in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl) using a Superose6™10/300 GL column at 0.4 ml/min flow-rate. B: BN-PAGE analysis of the elutions from SEC purification step of *Pab*MCM in a 4-12% Bis-Tris NativePAGE, gel was run for 3 hour at 150 V and de-stained o/n in 30% ethanol and 10% acetic acid. C: Size exclusion chromatography trace of *Pab*MCM in complex with y-shaped DNA in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl) using a Superose6™10/300 GL column at 0.4 ml/min flow-rate. B: BN-PAGE analysis of the elutions from SEC purification step of *Pab*MCM in complex with y-shaped DNA in a 4-12% Bis-Tris NativePAGE, gel was run for 3 hour at 150 V and de-stained o/n in 30% ethanol and 10% acetic acid.

Figure 4.13 shows the SAXS profiles of *Pab*MCM in complex with different DNA substrates, in the absence or presence of ATPγS, which is known to stabilize the complex.

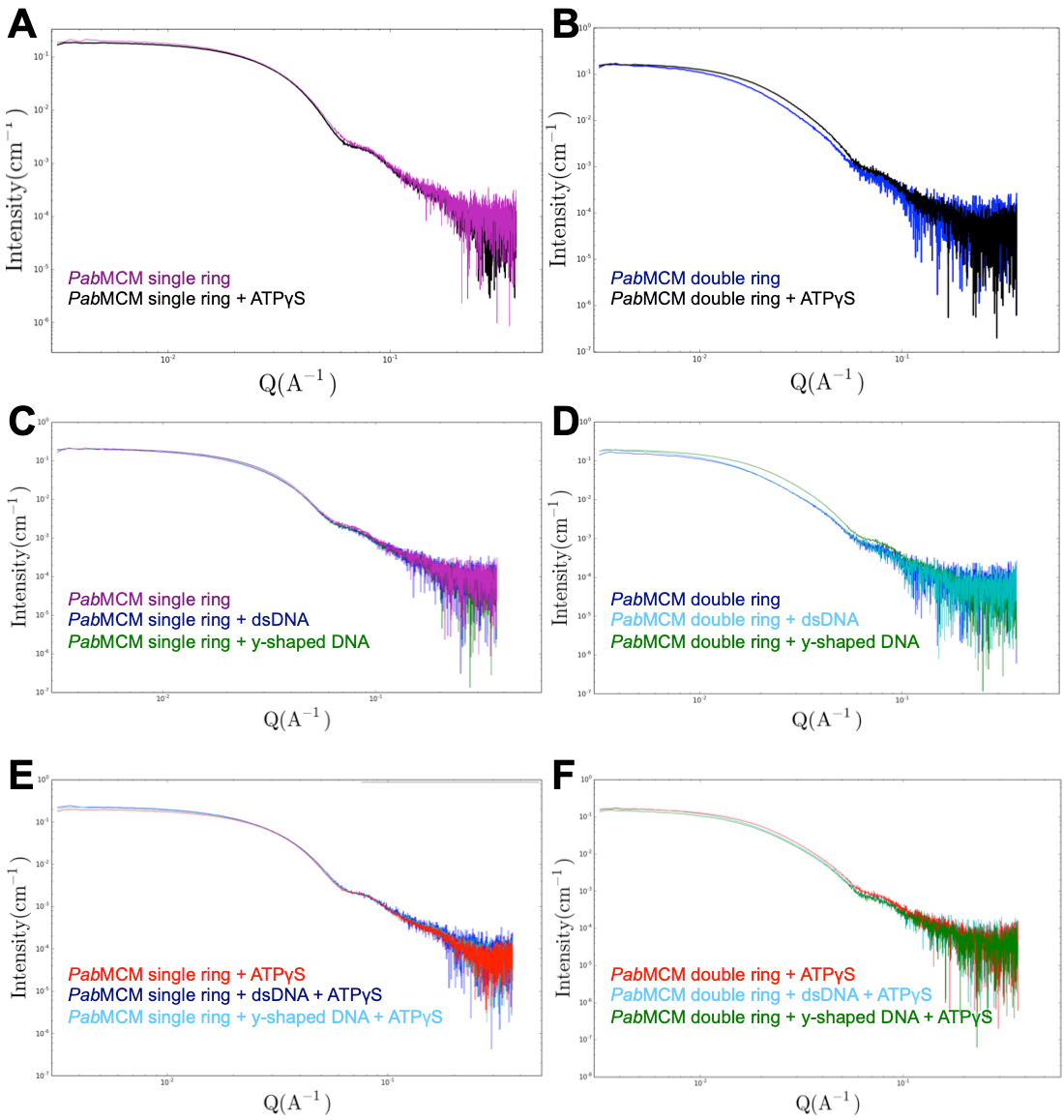
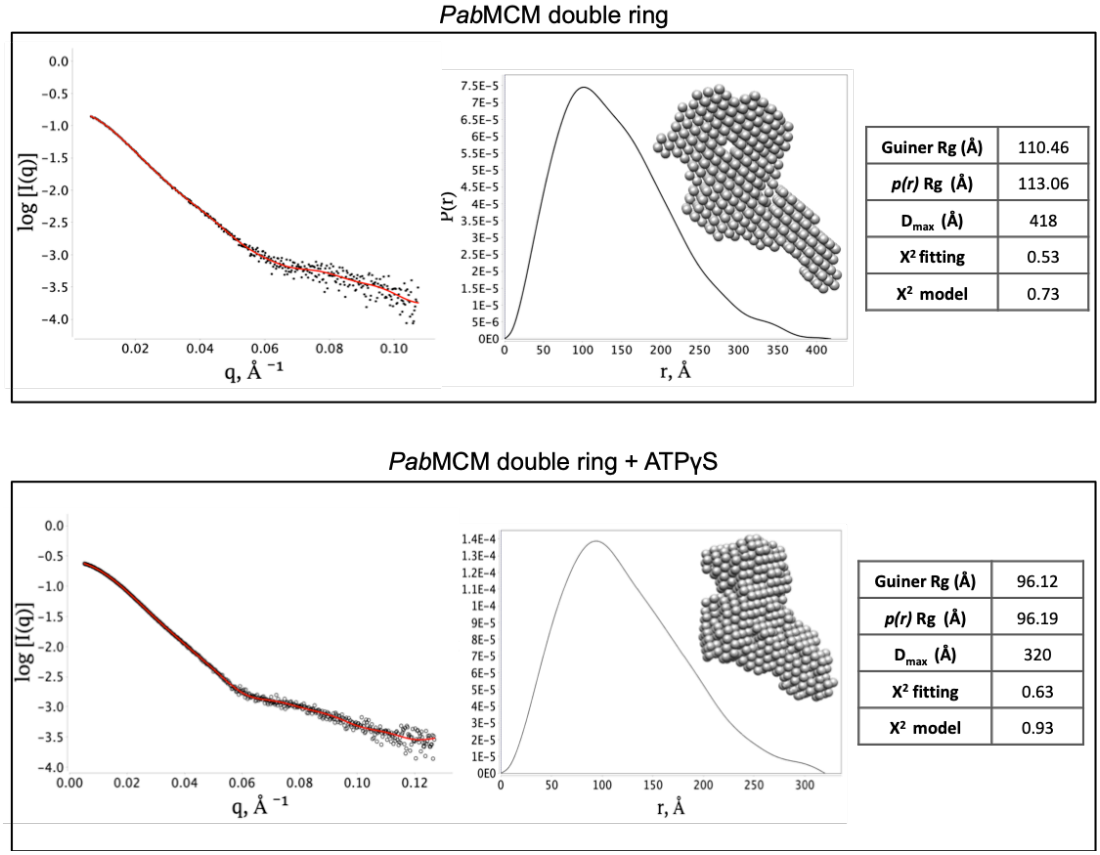


Figure 4.13: SAXS curves of *Pab*MCM in absence or presence of ATPγS and dsDNA or y-shaped DNA. A: Experimental SAXS intensity curves of *Pab*MCM in single ring assembly (purple) and in presence of ATPγS (black). B: Experimental SAXS intensity curves of *Pab*MCM in double ring assembly (blue) and in presence of ATPγS (black). C: Experimental SAXS intensity curves of *Pab*MCM in single ring assembly (purple) and in presence of dsDNA (blue) or y-shaped DNA (green). D: Experimental SAXS intensity curves of *Pab*MCM in double ring assembly (blue) and in presence of dsDNA (cyan) or y-shaped DNA (green). E: Experimental SAXS intensity curves of *Pab*MCM in single ring assembly in presence of ATPγS (red) and in presence of dsDNA (blue) or y-shaped DNA (cyan) and ATPγS. F: Experimental SAXS intensity curves of *Pab*MCM in double ring assembly in presence of ATPγS (red) and in presence of dsDNA (cyan) or y-shaped DNA (green) and ATPγS. Figures were prepared using SasView.

Both *Pab*MCM single and double rings adopt a more compact conformation when ATPγS is bound in the AAA+ domain. This effect is more dramatic for *Pab*MCM in double assembly and a change in overall dimensions of the complex can be observed (Figure 4.14). In fact, in presence of ATPγS, the R_g calculated from the Guinier analysis drops from 110.46 Å to 96.12 Å (R_g [*p(r)*] drops from 113.06 Å to 96.19 Å), whereas the D_{max} drops from 418 Å to 320 Å. In both cases, the X^2 of the fitting are appropriate (0.53 vs 0.63) and the DAMMIN models obtained show an increased compactness and, consequently, a reduction of "extra density" when ATPγS is bound to the double ring.

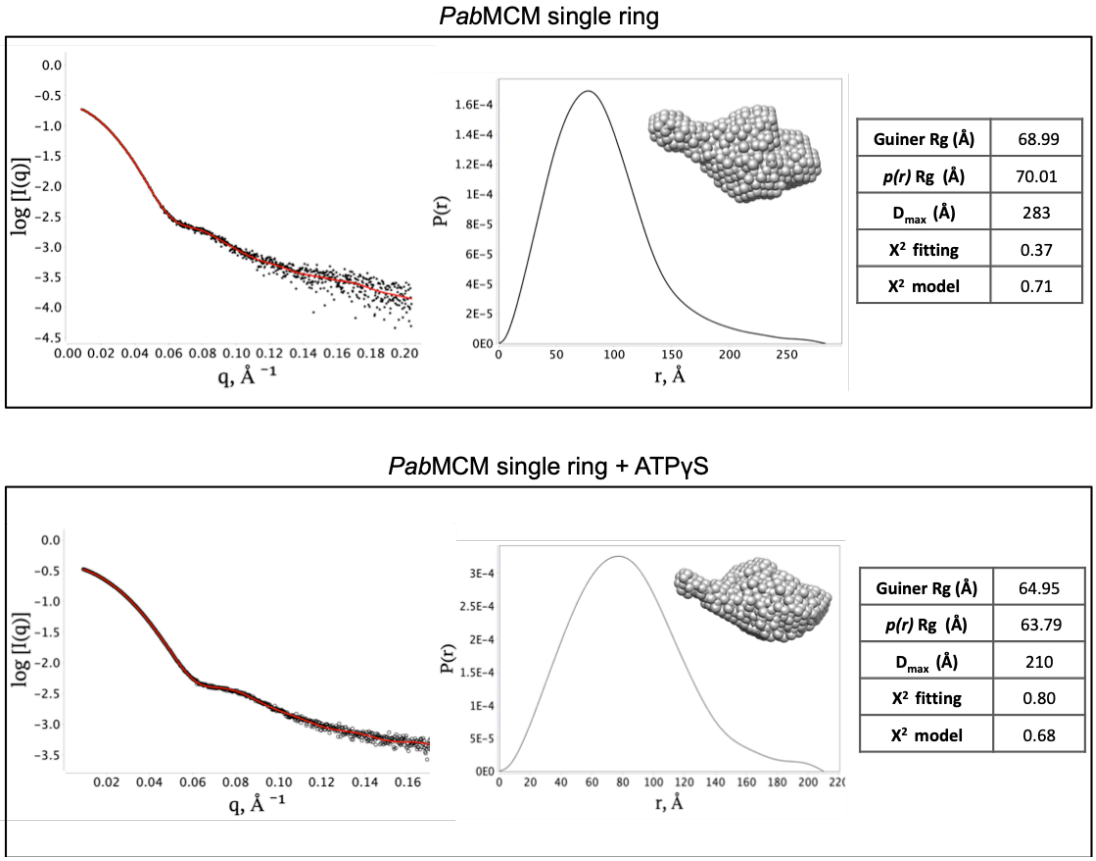


Guiner Rg (Å)	110.46
p(r) Rg (Å)	113.06
D_{max} (Å)	418
X^2 fitting	0.53
X^2 model	0.73

Guiner Rg (Å)	96.12
p(r) Rg (Å)	96.19
D_{max} (Å)	320
X^2 fitting	0.63
X^2 model	0.93

Figure 4.14: SAXS analysis of *Pab*MCM double ring in absence or presence of ATPγS. SAXS intensity curve, real space fitting (in red), *p(r)* distribution, DAMMIN model and table with overall dimensions of the complex and statistical analysis. Beads size of the models increased post-analysis for clarity in figure.

For what concerns for *Pab*MCM in single ring assembly a smaller change in overall dimensions of the complex can be observed (Figure 4.15). In fact, in presence of ATPγS, the R_g calculated from the Guinier analysis drops from 68.99 Å to 64.95 Å (R_g [*p(r)*] drops from 70.01 Å to 63.79 Å), whereas the D_{max} drops from 283 Å to 210 Å. In both cases, the X^2 of the fitting are appropriate (0.37 vs 0.8) and the DAMMIN models obtained also show a reduction of extra density when ATPγS is bound to the single ring.



Guiner Rg (Å)	68.99
p(r) Rg (Å)	70.01
D_{max} (Å)	283
X^2 fitting	0.37
X^2 model	0.71

Guiner Rg (Å)	64.95
p(r) Rg (Å)	63.79
D_{max} (Å)	210
X^2 fitting	0.80
X^2 model	0.68

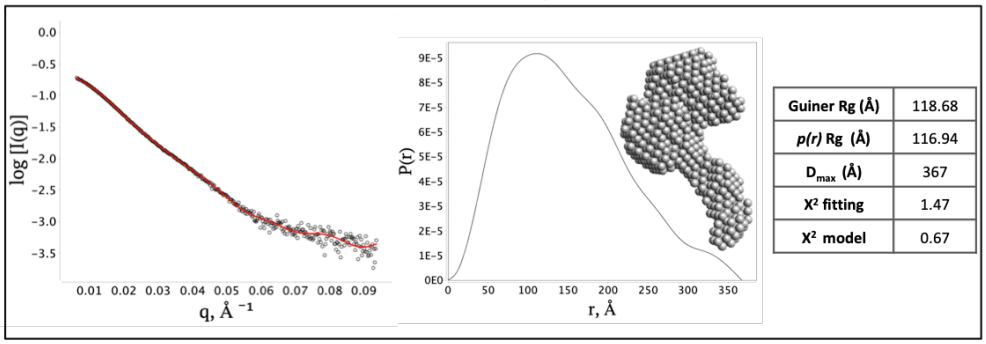
Figure 4.15: SAXS analysis of *Pab*MCM single ring in absence or presence of ATPγS. SAXS intensity curve, real space fitting (in red), *p(r)* distribution, DAMMIN model and table with overall dimensions of the complex and statistical analysis. Beads size of the models increased post-analysis for clarity in figure.

Similarly, when *Pab*MCM is bound to DNA substrates, the change induced by the binding of ATPγS is bigger for double ring assemblies (Figure 4.16) than for single rings (Figure 4.17). Moreover, the binding of dsDNA results in an increase of the R_g of the complex and in a decrease of its D_{max}, while the binding of y-shaped DNA results in a overall more compact assembly, comparable to the one induced by the binding of ATPγS.

More specifically, when *Pab*MCM double ring binds dsDNA, its R_g changes from 110.46 Å to 118.68 Å (R_g [*p(r)*] increases from 113.06 Å to 116.94 Å), whereas the D_{max} drops from 418 Å to 367 Å. Also in this case, the binding of ATPγS triggers the complex to adopt a more compact conformation and its R_g drops from 118.68 Å to 107.02 Å (R_g [*p(r)*] drops from 116.94 Å to 107.76 Å), whereas the D_{max} decreases from 367 Å to 357 Å. In addition, when *Pab*MCM double ring binds y-shaped DNA, its R_g changes from 110.46 Å to 96.84 Å (R_g [*p(r)*] drops from 113.06 Å to 98.43 Å), whereas the D_{max} drops from 418 Å to 356 Å. However, in this case, the binding of ATPγS does not induce a more compact conformation of the complex and, instead, its R_g increases from 96.84 Å to 104.50 Å (R_g [*p(r)*] increases from 98.43 Å to 103.86 Å) and the D_{max} slightly decreases from 356 Å to 352 Å.

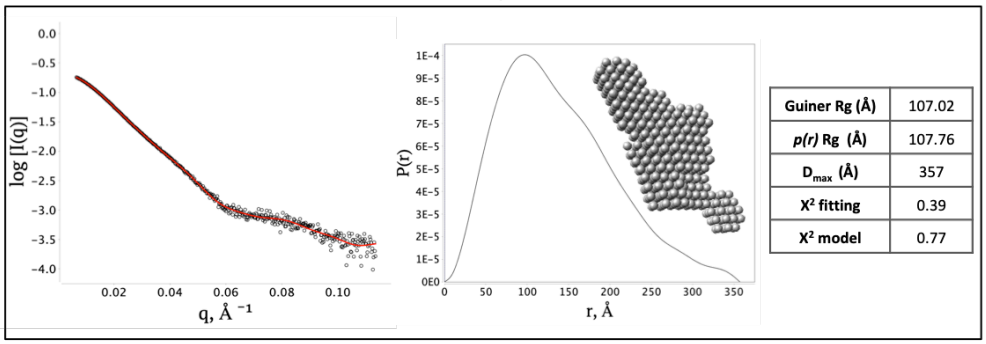
On the other hand, when *Pab*MCM in single ring assembly is bound to DNA substrates, an extremely small change in overall dimensions of the complex can be observed. In fact, in presence of dsDNA, the R_g calculated from the Guinier analysis slightly increases from 68.99 Å to 72.57 Å (R_g [*p(r)*] changes from 70.01 Å to 73 Å), whereas the D_{max} decreases from 283 Å to 255 Å. As expected, the binding of ATPγS induces a more compact conformation of the complex and its R_g calculated from the Guinier analysis slightly decreases from 72.57 Å to 70.85 Å (R_g [*p(r)*] changes from 73 Å to 70.05 Å), whereas the D_{max} drops from 255 Å to 240 Å. Additionally, when *Pab*MCM in single ring assembly is bound to y-shaped DNA, the R_g calculated from the Guinier analysis slightly increases from 68.99 Å to 70.12 Å (R_g [*p(r)*] changes from 70.01 Å to 71.07 Å), whereas the D_{max} decreases from 283 Å to 253 Å. Also in this case, as previously observed for *Pab*MCM in complex with y-shaped DNA, the binding of ATPγS does not induce a change of the overall dimensions of the complex and its R_g calculated from the Guinier analysis is 70.8 Å (R_g [*p(r)*] = 69.64 Å), while its D_{max} is 249 Å.

*Pab*MCM double ring + dsDNA



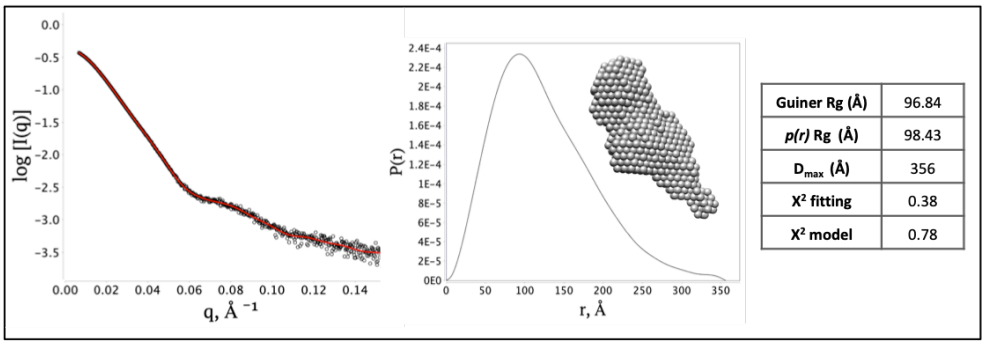
Guiner Rg (Å)	118.68
p(r) Rg (Å)	116.94
D_{max} (Å)	367
X^2 fitting	1.47
X^2 model	0.67

*Pab*MCM double ring + dsDNA + ATPγS



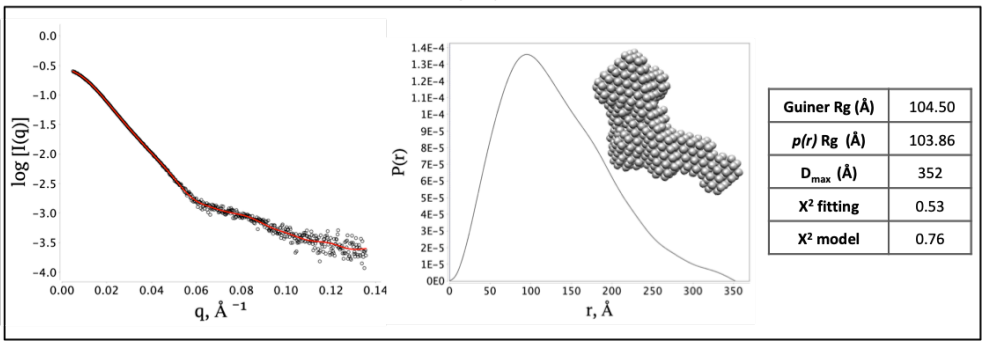
Guiner Rg (Å)	107.02
p(r) Rg (Å)	107.76
D_{max} (Å)	357
X^2 fitting	0.39
X^2 model	0.77

*Pab*MCM double ring + y-shaped DNA



Guiner Rg (Å)	96.84
p(r) Rg (Å)	98.43
D_{max} (Å)	356
X^2 fitting	0.38
X^2 model	0.78

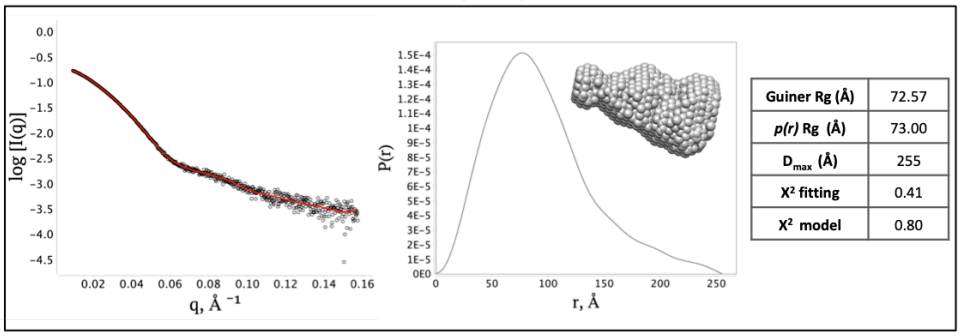
*Pab*MCM double ring + y-shaped DNA + ATPγS



Guiner Rg (Å)	104.50
p(r) Rg (Å)	103.86
D_{max} (Å)	352
X^2 fitting	0.53
X^2 model	0.76

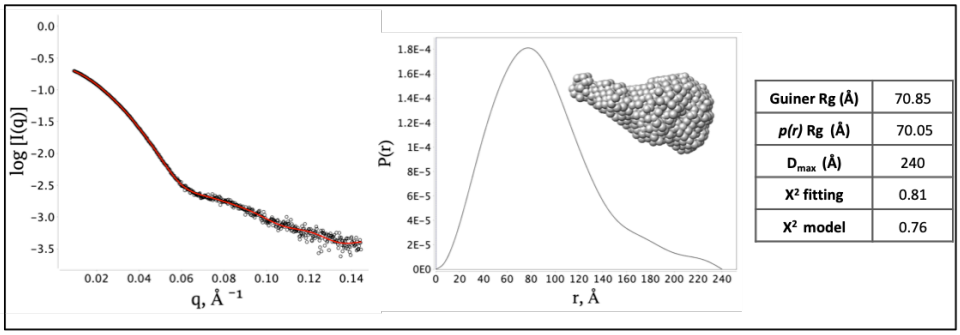
Figure 4.16: SAXS analysis of *Pab*MCM double ring in presence of DNA substrates and in absence or presence of ATPγS. SAXS intensity curve, real space fitting (in red), *p(r)* distribution, DAMMIN model and table with overall dimensions of the complex and statistical analysis. Beadsize of the models increased post-analysis for clarity in figure.

*Pab*MCM single ring + dsDNA



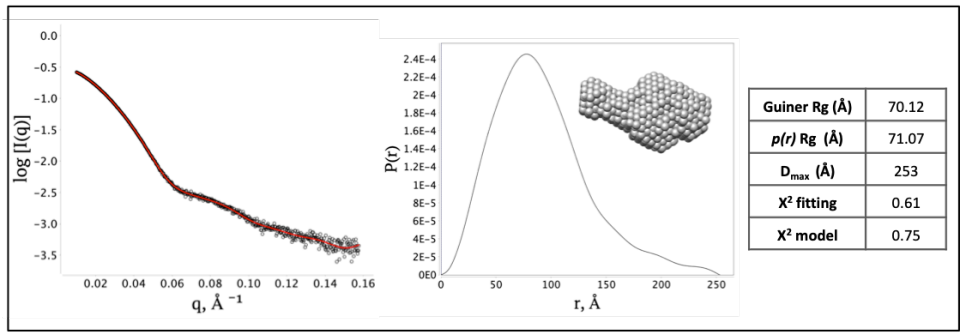
Guiner Rg (Å)	72.57
p(r) Rg (Å)	73.00
D_{max} (Å)	255
X^2 fitting	0.41
X^2 model	0.80

*Pab*MCM single ring + dsDNA + ATPγS



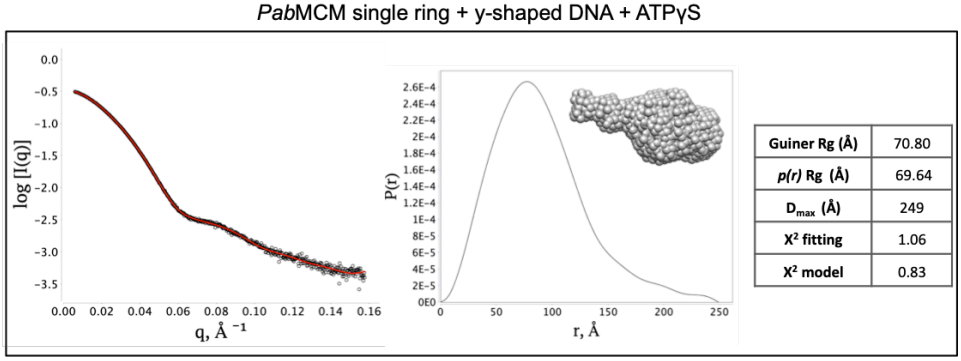
Guiner Rg (Å)	70.85
p(r) Rg (Å)	70.05
D_{max} (Å)	240
X^2 fitting	0.81
X^2 model	0.76

*Pab*MCM single ring + y-shaped DNA



Guiner Rg (Å)	70.12
p(r) Rg (Å)	71.07
D_{max} (Å)	253
X^2 fitting	0.61
X^2 model	0.75

*Pab*MCM single ring + y-shaped DNA + ATPγS



Guiner Rg (Å)	70.80
p(r) Rg (Å)	69.64
D_{max} (Å)	249
X^2 fitting	1.06
X^2 model	0.83

Figure 4.17: SAXS analysis of *Pab*MCM single ring in presence of DNA substrates and in absence or presence of ATPγS. SAXS intensity curve, real space fitting (in red), *p(r)* distribution, DAMMIN model and table with overall dimensions of the complex and statistical analysis. Bead size of the models increased post-analysis for clarity in figure.

4.3.5 Cryo-EM of *Pab*MCM in complex with y-shaped DNA and ATPγS

The architecture of *Pab*MCM in complex with y-shaped DNA and ATPγS was studied by cryo-EM coupled to single particle analysis in order to assess conformational heterogeneity. 500 µl of nickel purified *Pab*MCM protein (10 mg/ml) were incubated with 100 µl of y-shaped DNA (50 µM) and 5 mM ATPγS for 1 hour at room temperature, prior to size exclusion chromatography using a Superose 6 Increase 10/300 GL column (Figure 4.18A). Fractions from the gel filtration were screened by negative staining TEM in order to identify the fraction containing a similar proportion of single and double ring assemblies, which resulted to be fraction 12 (Figure 4.18B).

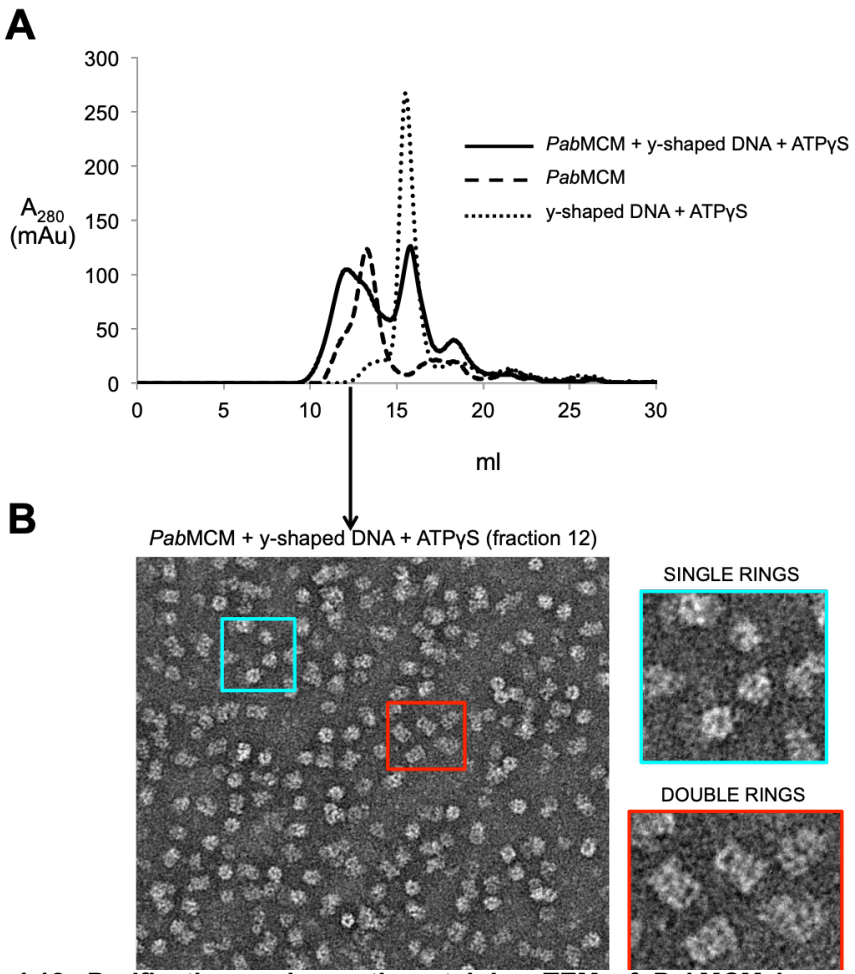


Figure 4.18: Purification and negative staining TEM of *Pab*MCM in complex with y-shaped DNA and ATPγS. A: SEC trace of *Pab*MCM+y-shaped DNA+ATPγS (continuous line), *Pab*MCM (dashed line), and y-shaped DNA+ATPγS (dotted line) using a Superose 6 Increase 10/300 GL column in SEC buffer. B: Negative staining TEM of fraction 12 of SEC of PabMCM+y-shaped DNA+ATPγS on a 400 mesh copper grid, using 2% uranyl acetate, where the cyan square shows single ring assemblies and the red square shows double ring assemblies.

The cryo-EM data collection of fraction 12 of the SEC of *Pab*MCM+y-shaped DNA+ATPγS was carried out at eBIC (Diamond Light Source, Harwell) under low-dose mode at nominal magnification of 50,000x, at a final sampling of 1.06 Å/pixel, using a FEI Titan Krios microscope, equipped with a Gatan Quantum K3 Summit image filter/detector. Data was processed using RELION software by myself and Dr Laura Spagnolo, independently. The data processing pipeline, from manual picking to 3D classification, presented in this section derives from my single particle analysis. Higher-resolution refined 3D maps were generated and provided by Dr Laura Spagnolo. Further analysis on the cryo-EM density maps, such as superposition of the subunits, were performed by myself. From the CTF and motion corrected dataset (~2500 micrographs) (Figure 4.19A), ~3,000 single particles (SP) were manually picked using 250 Å masks; the single particle images were extracted into boxes 400x400 pixels and subjected to one round of 2D classification in order to generate reference images for auto-picking. ~543,000 particles were autopicked, extracted into boxes 400x400 pixels and rescaled to 100x100 pixels. The dataset was subjected to sequential rounds of alignments and 2D classification using a 300 Å mask, in order to initially clean up the dataset and then to improve the resulting class-average images. The polished dataset (~390,000 SP) was divided into sub-datasets on the basis of the class-averages obtained with the 2D classification. Two sub-datasets were created for *Pab*MCM in single ring conformations (~206,000 SP) and in double ring conformation (~201,000 SP) (Figures 4.19B and 4.19C, respectively). Both the sub-datasets contain particles from the same top/bottom 2D class averages, while particles from 2D class averages that represent side views are sorted in the two sub-datasets. Initial models for the single and double ring complexes were generated with EMAN2 without imposing any symmetry and checked by inspecting the 3D model and the correspondence between the class averages and the re-projections of the models. Even though the addition of the y-shaped DNA seemed to decrease the level of asymmetry of the complex, we decided to avoid imposing symmetries at any stage of the analysis. An initial 3D classification was performed for each sub-dataset by generating five 3D classes. Representative 3D models are shown in (Figures 4.19B and 4.19C). However, at this stage, the individual classes were not further processed to obtain refined high-resolution density maps, because of the high computational power that was required and limited resources.

The final models of single and double ring assemblies were provided by Dr Laura Spagnolo (Figure 4.20).

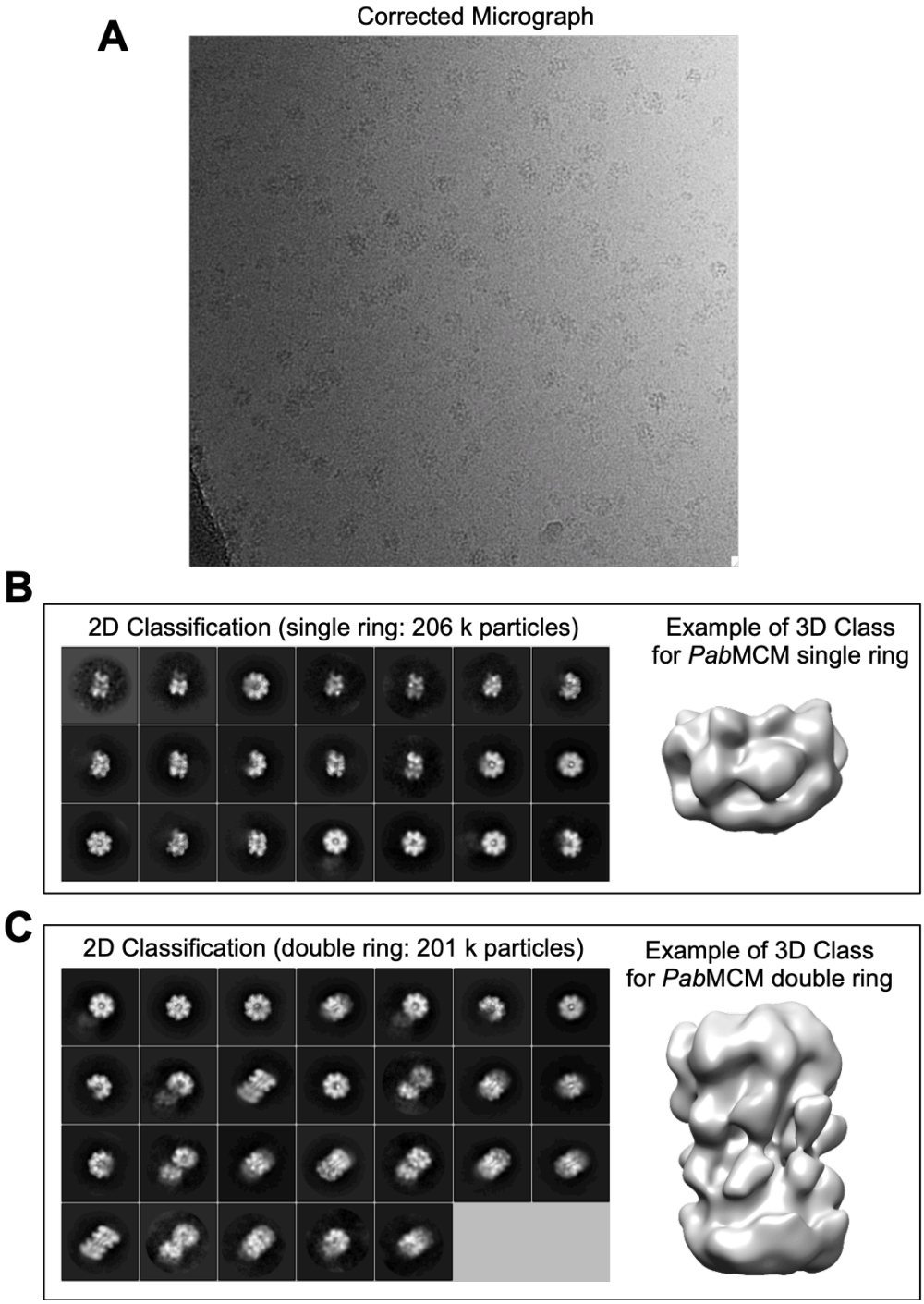


Figure 4.19: Cryo-EM of *Pab*MCM in complex with y-shaped DNA and ATPγS. A: Representative motion corrected micrograph of a cryo-EM dataset of fraction 12 of SEC of PabMCM+y-shaped DNA+ATPγS (low-dose mode, nominal magnification of 50,000x, final sampling of 1.06 Å/pixel, using a FEI Titan Krios microscope, equipped with a Gatan Quantum K3 Summit image filter/detector. B: 2D class averages of single ring assemblies and representative low-resolution 3D model. C: 2D class averages of double ring assemblies and representative low-resolution 3D model.

The cryo-EM density maps, obtained from single particle analysis of *Pab*MCM in complex with y-shaped DNA and ATPγS, show the presence of two distinct single ring assemblies and two different double ring assemblies (Figure 4.20). The architecture of *Pab*MCM, represented by the cryo-EM models obtained, will be discussed in more detail in the next paragraphs, with the exception of the one shown in Figure 4.20C. In fact, even though it appears clear that this model represents a double ring assembly where the two single rings are in closer contact, it is not possible to provide additional information, as the domain organization of the monomeric *Pab*MCM, as well as individual subunits of the complex, cannot be clearly detected.

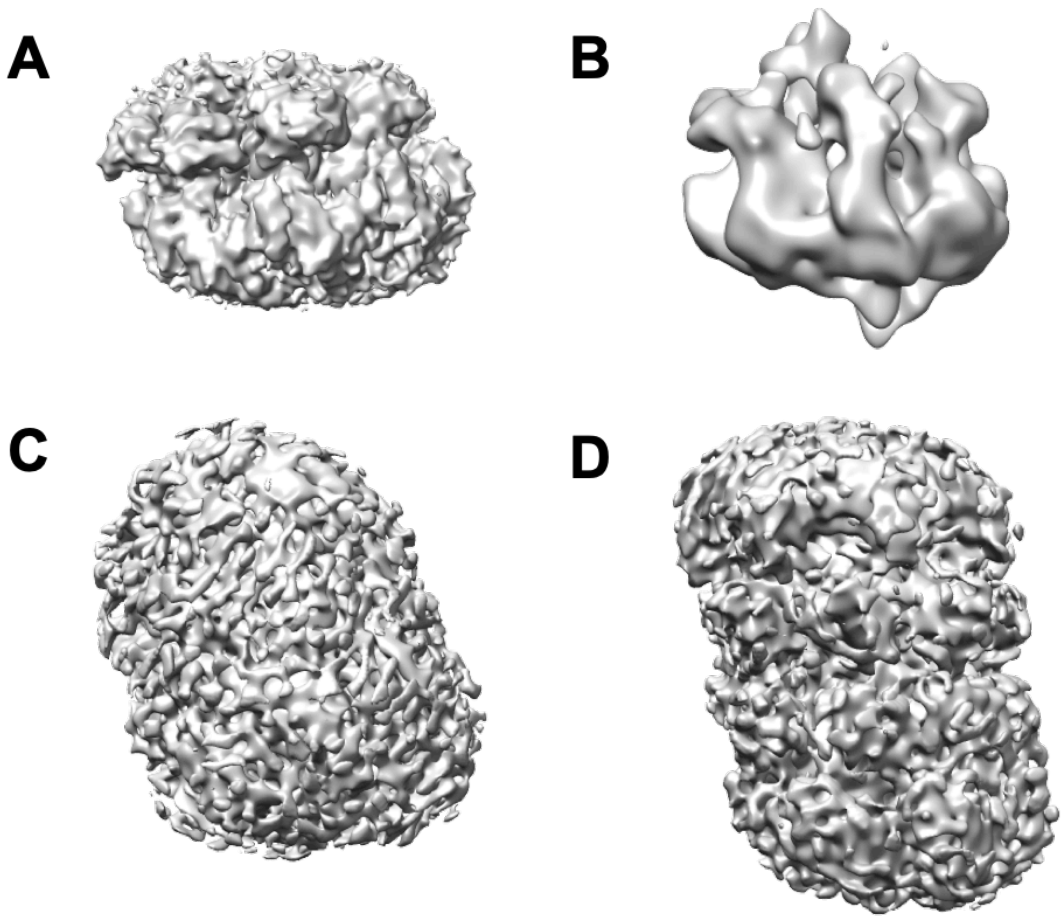


Figure 4.20: Cryo-EM models of *Pab*MCM in complex with DNA and ATPγS. A: 3D reconstruction of *Pab*MCM in single ring open assembly (~85,000 SP). B: 3D reconstruction of *Pab*MCM in single ring close assembly (~30,000 SP). C: 3D reconstruction of *Pab*MCM in double ring assembly (~25,000 SP). D: 3D reconstruction of *Pab*MCM in double ring assembly (~30,000 SP).

The high-resolution cryo-EM density map obtained for *Pab*MCM in single ring assembly show that the complex binds to y-shaped DNA as a hexamer, in an opened conformation. As shown in Figure 4.21A, a region of the y-shaped DNA is located in the middle of the complex but leans on one side of it, making contact with one of the subunits. Moreover, another part of the y-shaped DNA appears to twirl on the N-terminal part of the complex. The diameters of these DNA segments correspond to a single strand, for the region that can be found in the middle of the MCM complex, and to a double strand, for the one that sits on the surface of the ring where the N-terminal domains of each subunit make contact. Similarly, a recent study of the eukaryotic *E.cuniculi* MCM2-7, reporting the crystal structure of a near full-length MCM hexamer bound to ssDNA, shows that the open ring adopts a skewed conformation in order to accommodate the entrance of ssDNA through a spiral staircase (Meagher et al., 2019).

Figure 4.21B shows top and side views of the open hexameric ring bound to y-shaped DNA and the manual superposition of six identical copies of monomeric *Pab*MCM, obtained by homology modelling using Phyre2. The six subunits do not reside on the same plane, resulting in a skewed open ring. Consequently, the two external subunits are extremely different when the ring is observed from the top. In fact, "*Pab*MCM 1" appears to be smaller when compared to the other subunits, while "*Pab*MCM 6" appears bigger because of their relative position down the view axis. Moreover, it is possible to observe that both the N-terminal and the AAA+ domains of "*Pab*MCM 6" are visible from the top view. This is another consequence of the non-planar nature of the open ring and might be mistaken as indication of the presence of a seventh subunit, as previously shown in the 2D classification. Furthermore, in the cryo-EM map it is possible to detect a conformational change in the spatial organization of the winged helix-turn-helix domain of each *Pab*MCM subunit. In fact, the linker between the AAA+ and the C-terminal domains triggers a rearrangement of the latter towards the centre of the MCM complex. As previously described, the C-terminal domain of MCM might play a role in DNA-stimulated ATPase activity. In our model it is possible to detect points of contact between the C-terminal domains of two subunits and the DNA substrate and further analysis could provide a deeper understanding of the interaction between subunits and DNA.

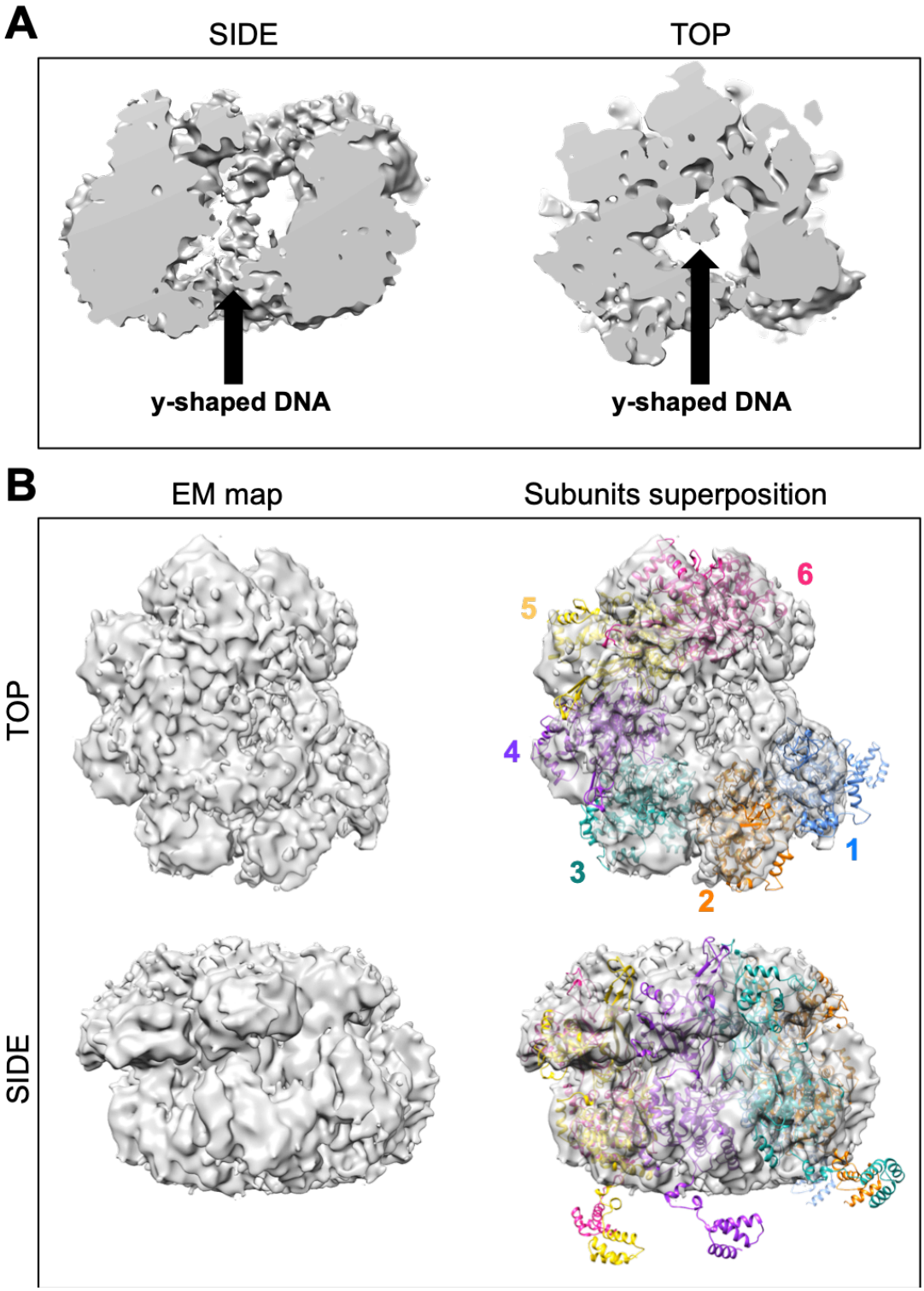


Figure 4.21: Cryo-EM model of hexameric *Pab*MCM single ring in complex with DNA and ATPγS. A: Side and top cut views of *Pab*MCM single ring highlighting the DNA path in the middle channel of the ring. B: Side and top views of *Pab*MCM single ring and superposition of six subunits of the homology model obtained for *Pab*MCM using Phyre2.

The low-resolution cryo-EM density map obtained for *Pab*MCM in single ring assembly shows that the complex can bind to y-shaped DNA as an heptamer. As shown in Figure 4.22A, the y-shaped DNA is located in the middle of the complex, leaning transversally. However, in this case, it appears that the double stranded region of the y-shaped DNA resides in the centre of the MCM complex, while the two single stranded segments stick out from the N-terminal side. Moreover, it is possible to observe that the N-terminal domains of several subunits mediate the contact with each single stranded portion of the y-shaped DNA, while some C-terminal domains seem to interact with the double stranded region.

Figure 4.22B shows top and side views of the heptameric ring bound to y-shaped DNA and the manual superposition of seven identical copies of monomeric *Pab*MCM. As described for the open MCM hexamer bound to y-shaped DNA, and also for the octameric MCM in absence of DNA reported in section 4.3.2 of this chapter, the subunits do not reside on the same plane. The resolution of this model does not allow superposition of individual subunits on the basis of recognizable secondary structure elements, but only on lower resolution densities that can be associated to N-terminal and AAA+ domains. The superposition of the N-terminal domains in this map suggests that they adopt a different orientation in the heptameric assembly, when compared to the one observed in the open hexamer. This is consistent with the fact that the N-terminal domain of MCM plays a key role in protein oligomerization and therefore, different oligomeric states require a different organization of the N-terminal domains of each subunit.

As described for the single hexameric ring in open conformation, the linker between the AAA+ and the C-terminal domains triggers a rearrangement of the latter towards the centre of the MCM complex. Interactions between the C-terminal domains and the DNA substrate can be observed.

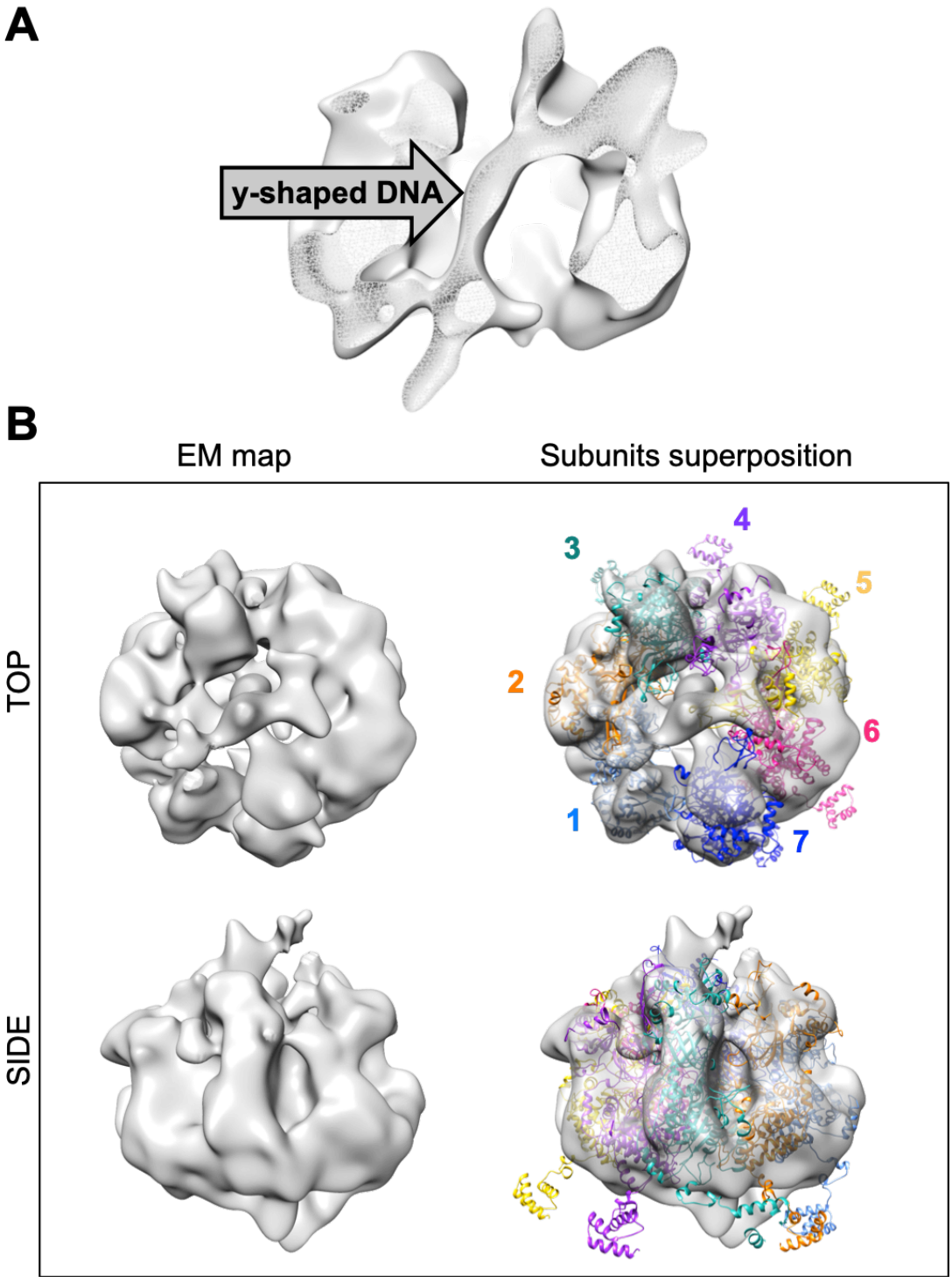


Figure 4.22: Cryo-EM model of heptameric *Pab*MCM single ring in complex with DNA and ATPγS. A: Side cut view of *Pab*MCM single ring highlighting the DNA path in the middle channel of the ring. B: Side and top views of *Pab*MCM single ring and superposition of seven subunits of the homology model obtained for *Pab*MCM using Phyre2.

The cryo-EM density map obtained for *Pab*MCM in double ring assembly shows that the complex can bind to DNA also as a double octameric assembly.

As shown in Figure 4.23A, the y-shaped DNA is buried inside the middle channel of the double ring complex. From our model, it appears that one of the *Pab*MCM rings mediates the contact with the double stranded region of the y-shaped DNA, while the other interacts with the single stranded segments. In fact, the “bottom” ring of the complex seems to accommodate one of the single stranded regions in its central channel, while the other appears to enter one of the side channels between subunits or between the N-terminal rings, as proposed in Brewster and Chen (2010). Two recent cryo-EM structures of the yeast MCM2-7 double hexamer bound to DNA shows that the two rings are staggered to accommodate a kinked dsDNA molecule of 60-62 bp in length (Noguchi et al., 2017, Abid Ali et al., 2017). In particular, the model proposed by Noguchi *et al.* suggests a lagging strand DNA extrusion model, where the DNA strand separation is caused by torsion of the N-tier rings, which results in the extrusion of the lagging strand from the centre of the double ring and, consequently, in the activation of the helicase.

Figure 4.23B shows top and side views of the double octameric ring in a head-to-head conformation, bound to y-shaped DNA. The manual superposition of eight identical copies of monomeric *Pab*MCM in each ring shows that the complex is not symmetrical, as all the subunits do not reside on the same plane. In fact, in our map it is possible to observe that the subunits 7, 8 and 9 of the two rings are more deeply interconnected through their N-terminal domains. This results in a more compact conformation of one side of the complex and therefore leads to the formation of an asymmetric complex.

As previously described for the single hexameric ring in open conformation and for the single octamer, the rearrangement of the linkers between the AAA+ and the C-terminal domains results in a re-positioning of the C-terminal domains towards the centre of the MCM complex, where they mediate interactions with the two extremities of the y-shaped DNA substrate, as described above.

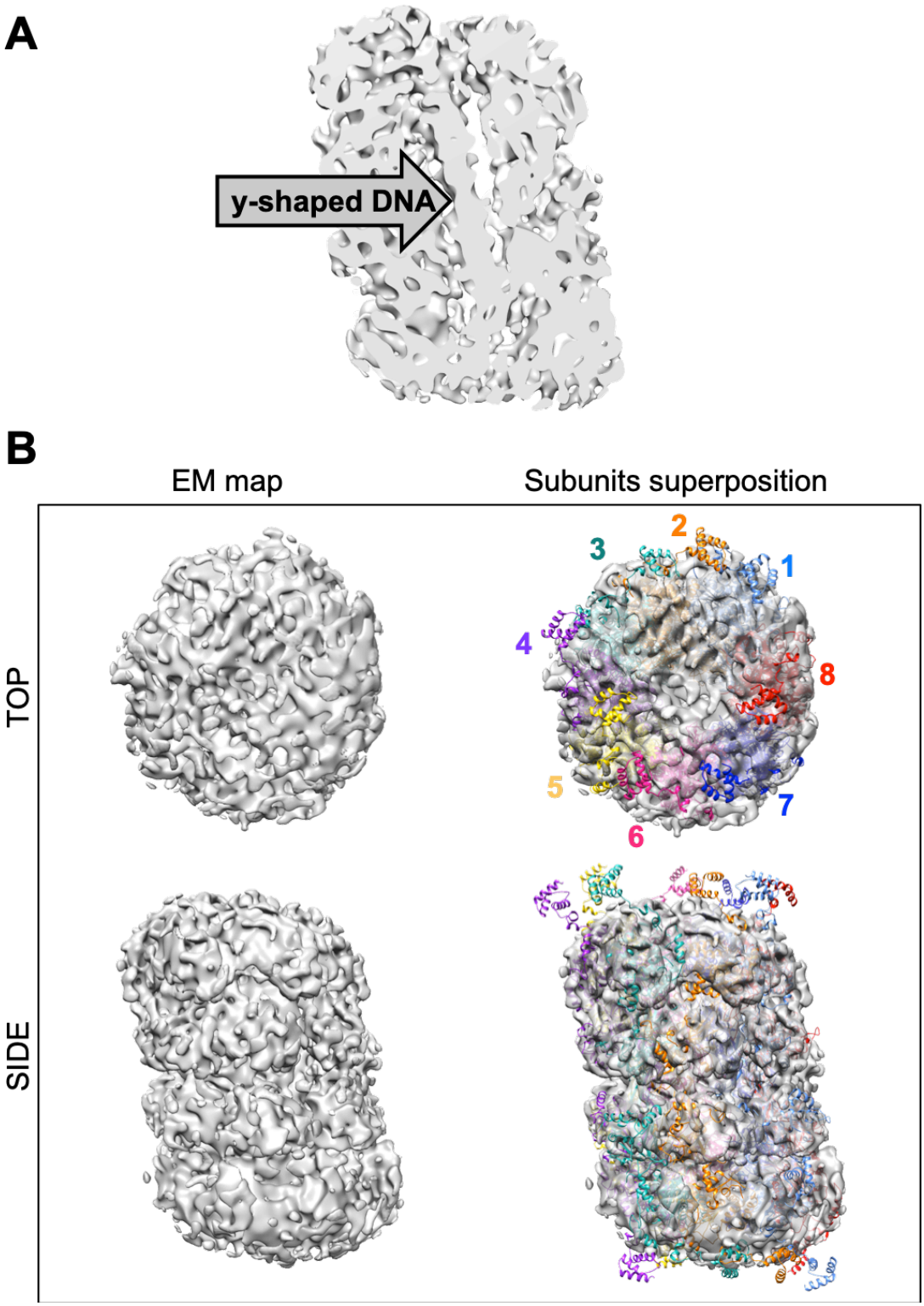


Figure 4.23: Cryo-EM model of octameric *Pab*MCM double ring in complex with DNA and ATPγS. A: Side cut view of *Pab*MCM double ring highlighting the DNA path in the middle channel of the ring. B: Side and top views of *Pab*MCM double ring and superposition of eight subunits of the homology model obtained for *Pab*MCM using Phyre2 for each ring, for a total of sixteen subunits.

4.4 Conclusions

Archaeal MCMs can form different assemblies *in vitro*, such as hexamers, heptamers, octamers -in single and double ring conformations- and filaments (Costa et al., 2006a, Costa et al., 2006b, Costa et al., 2008, Gomez-Llorente et al., 2005, Fletcher et al., 2003, Slaymaker et al., 2013, Cannone et al., 2017). Importantly, the question of whether hexamers, heptamers and octamers are all present in Archaeon cells and whether this polymorphism has a biological significance *in vivo*, still remain unsolved. With current knowledge, it is not possible to exclude that the presence of these diverse homo-multimers could be linked to artefacts driven by overexpression of these recombinant proteins and purification conditions.

Here, we used the *Pab*MCM protein as a functional tool to address advantages and limitations of different structural biology techniques, when dealing with a heterogeneous sample.

As shown in this study, as well as in Cannone, *et al.* (2017), *Pab*MCM binds DNA and has weak helicase activity *in vitro*. The limited activity of this protein could be linked to the absence of additional elements, such as GINS, Cdc6 and post-translational modifications, which are all known to enhance helicase activity in other thermophilic MCMs (Cannone et al., 2017, Pan et al., 2011, Yoshimochi et al., 2008, Ogino et al., 2014, Xia et al., 2015).

The polymorphism observed in archaeal MCM proteins adds an extra layer of complexity in the structural characterization of this complex. In particular, our cryo-EM analysis of *Pab*MCM, in isolation and bound to DNA, shows the presence of open hexamers, heptamers and octamers, both in single and double ring assemblies. Interestingly, only single ring and double ring species can be separated by size exclusion chromatography, suggesting that there might be an equilibrium between multimers within these two configurations. The dynamic behaviour of these complexes in solution was analysed by small angle X-ray scattering. However, the extreme complexity of this system, as well as the low-resolution information that can be obtained with this technique, did not allow a quantitative measure of the heterogeneity observed. On the other hand, using small angle scattering methods (X-ray and neutron), we were able to detect a

change in overall dimension of the *Pab*MCM complex when bound to y-shaped DNA, but not additional information on where the DNA is bound and on the conformation of the complex. Therefore, particular attention should be taken when interpreting these results. In fact, the difference in dimensions is likely to derive from an increase in the number of double ring assemblies, which results in augmented heterogeneity that might be difficult to detect. Nevertheless, these analyses allowed us to choose a y-shaped DNA molecule for further cryo-EM studies on *Pab*MCM. Despite the different oligomeric state of *Pab*MCM complexes, our cryo-EM data are in agreement with the models proposed by recent studies on the eukaryotic MCM2-7 complexes bound to different DNA molecules (Noguchi et al., 2017, Abid Ali et al., 2017, Meagher et al., 2019). In particular, our single hexameric ring in open conformation bound to y-shaped DNA resembles the recent crystal structure of a near full-length *E.cuniculi* MCM2-7 hexamer bound to ssDNA. The mechanism proposed by this study is that the open ring adopts a skewed conformation in order to accommodate the entrance of ssDNA through a spiral staircase (Meagher et al., 2019). Similarly, we suggest that an open hexameric *Pab*MCM ring can wrap single stranded DNA at the melted origin of replication and proceed towards the dsDNA to unwinding it, excluding the lagging strand from its N-tier ring (Figure 4.24).

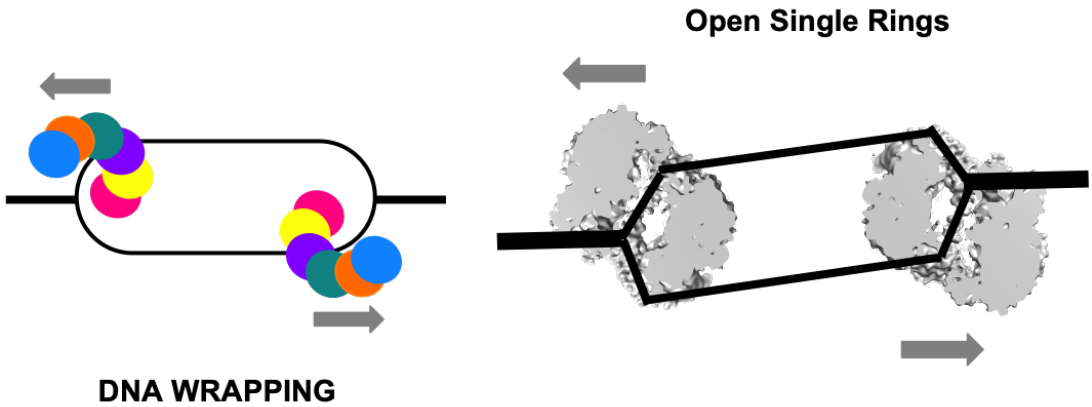


Figure 4.24: Model of the mechanism of action of *Pab*MCM in ssDNA wrapping at a melted origin of replication and dsDNA unwinding.

Additionally, our double octameric ring bound to y-shaped DNA resembles the recent cryo-EM maps obtained for yeast MCM2-7 double hexamer bound to dsDNA. The mechanism proposed by these studies suggests a lagging strand DNA extrusion model, where the DNA strands separation is caused by torsion of the N-tier rings, which results in extrusion of the lagging strand from the centre of the double ring and, consequently, in the activation of the helicase.

Similarly, we suggest that a double octameric ring *Pab*MCM complex can wrap double stranded DNA at the origin of replication and the tilted torsion of the two rings causes the DNA strands to separate, with one of them remaining in the central channel and the other one be extruded from the centre of the double ring. Upon activation, the two single octameric rings can proceed in 3'→5' direction unwinding dsDNA (Figure 4.25).

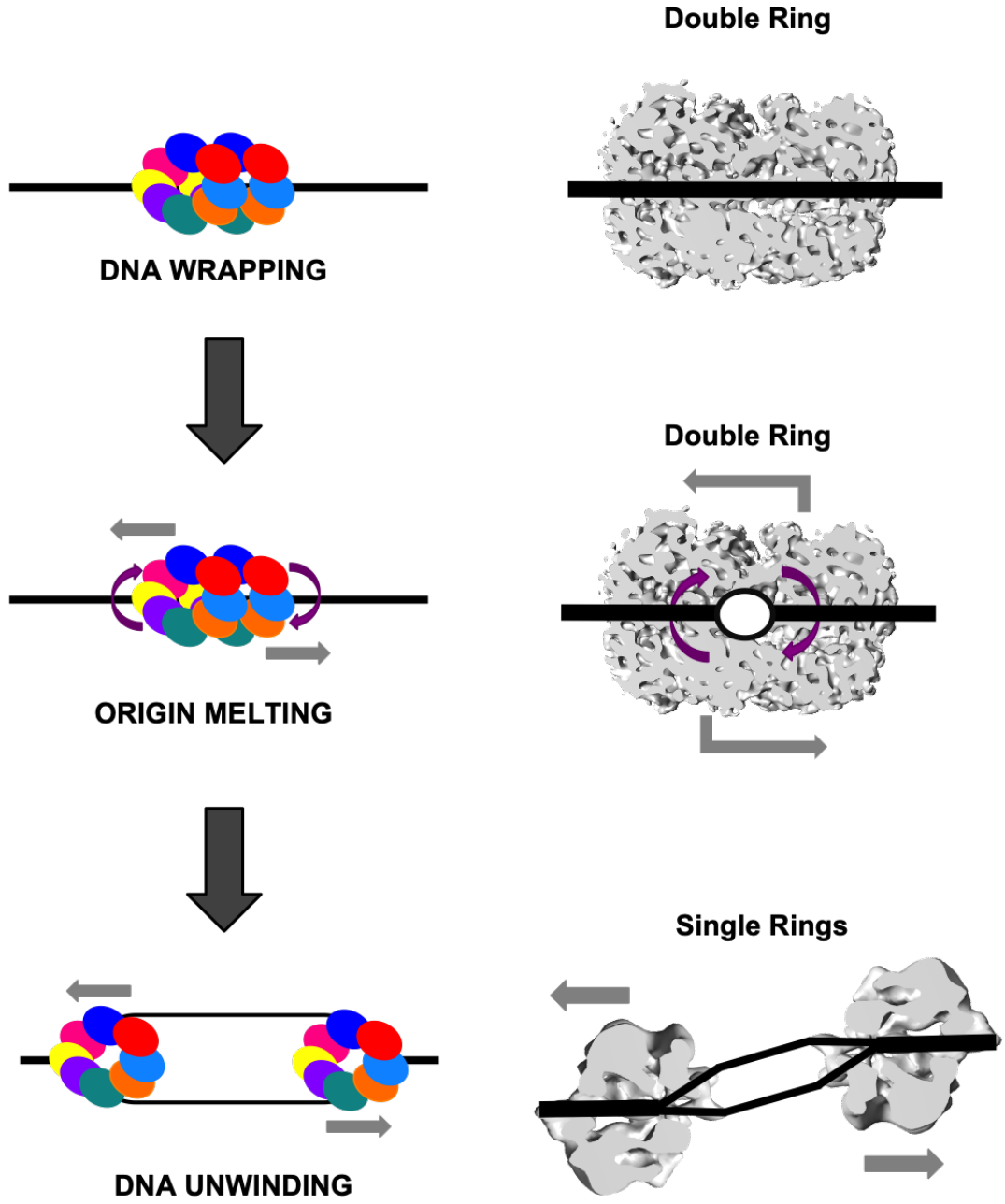


Figure 4.25: Model of the mechanism of action of *Pab*MCM in DNA replication, from dsDNA wrapping at the origin of replication, to origin melting and DNA unwinding.

Chapter 5

Conclusions and future perspectives

In this thesis, I showed how an integrated structural biology approach can be used to study two extremely different nucleic acids binding proteins, the human Staufen1 protein and the archaeal *Pab*MCM complex.

These two proteins are examples of complicated systems, which cannot be fully characterized only by adopting one technique. In fact, the structural characterization of non-ideally behaving systems requires a multi-angle approach that provides information at different levels of detail. Adopting an integrative structural biology approach, by combining several biochemical, bioinformatical and biophysical techniques, allows to obtain an all-round model of the system of interest.

In particular, the combination of hydrodynamic methods, homology modelling, small angle X-ray scattering and NMR allowed us to demonstrate that Staufen1 is a highly flexible protein, that recoils in solution from an elongated to a compact conformation in which the domains are in closer proximity but not interacting with each other. The information gathered using this multi-angle approach allowed us to propose a model in which flexible inter-domain loops possess a regulatory role in Staufen activity. We suggest that the high degree of flexibility of this protein could be functionally linked to the involvement of Staufen in many diverse aspects of RNA metabolism and regulation, by allowing a wide range of interactions with many RNA targets and protein partners.

Additionally, the higher degree of polymorphism observed in *Pab*MCM, compared to other archaeal and eukaryotic MCM proteins, adds an extra layer of complexity in the structural characterization of this complex. In this case, the use of small angle scattering could only provide a way to assess dynamic changes in overall dimensions of the assemblies and cryo-EM represents the best resource available to date to structurally address the polymorphism of archaeal MCM proteins and gain insights on its functional relevance in DNA replication.

The recent advances in electron microscopy led to a “resolution revolution” resulting with the establishment of this technique as a tool to obtain near-atomic resolution structures, previously achieved only by X-ray crystallography or NMR spectroscopy. Moreover, the development of the phase plate allowed to determine the architecture of small proteins (<100 kDa), which were not traditional targets of cryo-EM studies. In fact, we were able to directly visualise the monomeric Staufen1 protein, which is only 55 kDa and just below the limit of size of proteins that have been reported to be studied with cryo-EM. Additionally, cryo-EM is the tool that allows to study the structure of heterogeneous systems, such as archaeal MCM proteins. Advances in the cryo-TEM (tomography) field are making possible to visualize single enzymes inside the cells and this could allow us to address the question of the functional relevance of polymorphism in archaeal proteins *in vivo*. Finally, continuous and rapid improvements in direct detectors, phase plates and computational methods are likely to break many of the barriers that have been encountered in the structural characterization of problematic and unconventional systems.

Bibliography

ABID ALI, F., DOUGLAS, M. E., LOCKE, J., PYE, V. E., NANS, A., DIFFLEY, J. F. X. & COSTA, A. 2017. Cryo-EM structure of a licensed DNA replication origin. *Nat Commun,* 8**,** 2241.

ALLISON, R., CZAPLINSKI, K., GIT, A., ADEGBENRO, E., STENNARD, F., HOULISTON, E. & STANDART, N. 2004. Two distinct Staufen isoforms in Xenopus are vegetally localized during oogenesis. *RNA,* 10**,** 1751-63.

APARICIO, T., GUILLOU, E., COLOMA, J., MONTOYA, G. & MENDEZ, J. 2009. The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res,* 37**,** 2087-95.

APARICIO, T., IBARRA, A. & MENDEZ, J. 2006. Cdc45-MCM-GINS, a new power player for DNA replication. *Cell Div,* 1**,** 18.

BAE, B., CHEN, Y. H., COSTA, A., ONESTI, S., BRUNZELLE, J. S., LIN, Y., CANN, I. K. & NAIR, S. K. 2009. Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure,* 17**,** 211-22.

BAILIS, J. M. & FORSBURG, S. L. 2004. MCM proteins: DNA damage, mutagenesis and repair. *Curr Opin Genet Dev,* 14**,** 17-21.

BAILIS, J. M., LUCHE, D. D., HUNTER, T. & FORSBURG, S. L. 2008. Minichromosome maintenance proteins interact with checkpoint and recombination proteins to promote s-phase genome stability. *Mol Cell Biol,* 28**,** 1724-38.

BALACHANDRAN, S. & BARBER, G. N. 2007. PKR in innate immunity, cancer, and viral oncolysis. *Methods Mol Biol,* 383**,** 277-301.

BALTZ, A. G., MUNSCHAUER, M., SCHWANHAUSSER, B., VASILE, A., MURAKAWA, Y., SCHUELER, M., YOUNGS, N., PENFOLD-BROWN, D., DREW, K., MILEK, M., WYLER, E., BONNEAU, R., SELBACH, M., DIETERICH, C. & LANDTHALER, M. 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell,* 46**,** 674-90.

BANERJEE, A., BENJAMIN, R., BALAKRISHNAN, K., GHOSH, P. & BANERJEE, S. 2014. Human protein Staufen-2 promotes HIV-1 proliferation by positively regulating RNA export activity of viral protein Rev. *Retrovirology,* 11**,** 18.

BARRAUD, P. & ALLAIN, F. H. T. 2012. ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Current topics in microbiology and immunology,* 353**,** 35-60.

BARRY, E. R., MCGEOCH, A. T., KELMAN, Z. & BELL, S. D. 2007. Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic Acids Res,* 35**,** 988-98.

BECKMANN, B. M., CASTELLO, A. & MEDENBACH, J. 2016. The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Arch,* 468**,** 1029-40.

BECKMANN, B. M., HOROS, R., FISCHER, B., CASTELLO, A., EICHELBAUM, K., ALLEAUME, A. M., SCHWARZL, T., CURK, T., FOEHR, S., HUBER, W., KRIJGSVELD, J. & HENTZE, M. W. 2015. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun,* 6**,** 10127.

BÉLANGER, G., STOCKSLEY, M. A., VANDROMME, M., SCHAEFFER, L., FURIC, L., DESGROSEILLERS, L. & JASMIN, B. J. 2003. Localization of the RNA-binding proteins Staufen1 and Staufen2 at the mammalian neuromuscular junction. *Journal of Neurochemistry,* 86**,** 669-677.

BLACKHAM, S. L. & MCGARVEY, M. J. 2013. A host cell RNA-binding protein, Staufen1, has a role in hepatitis C virus replication before virus assembly. *J Gen Virol,* 94**,** 2429-36.

BOCHMAN, M. L. & SCHWACHA, A. 2010. The Saccharomyces cerevisiae Mcm6/2 and Mcm5/3 ATPase active sites contribute to the function of the putative Mcm2-7 'gate'. *Nucleic Acids Res,* 38**,** 6078-88.

BONNET-MAGNAVAL, F., PHILIPPE, C., VAN DEN BERGHE, L., PRATS, H., TOURIOL, C. & LACAZETTE, E. 2016. Hypoxia and ER stress promote Staufen1 expression through an alternative translation mechanism. *Biochem Biophys Res Commun,* 479**,** 365-371.

BOULAY, K., GHRAM, M., VIRANAICKEN, W., TREPANIER, V., MOLLET, S., FRECHINA, C. & DESGROSEILLERS, L. 2014. Cell cycle-dependent regulation of the RNA-binding protein Staufen1. *Nucleic Acids Res,* 42**,** 7867-83.

BRENDZA, R. P., SERBUS, L. R., DUFFY, J. B. & SAXTON, W. M. 2000. A Function for Kinesin I in the Posterior Transport of oskar mRNA and Staufen Protein. *Science,* 289**,** 2120-2122.

BREWSTER, A. S. & CHEN, X. S. 2010. Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. *Crit Rev Biochem Mol Biol,* 45**,** 243-56.

BREWSTER, A. S., WANG, G., YU, X., GREENLEAF, W. B., CARAZO, J. M., TJAJADI, M., KLEIN, M. G. & CHEN, X. S. 2008. Crystal structure of a near-full-length archaeal MCM: Functional insights for an AAA hexameric helicase. *PNAS,* 105**,** 20191-20196.

BUCHNER, G., BASSI, M. T., ANDOLFI, G., BALLABIO, A. & FRANCO, B. 1999. Identification of a novel homolog of the Drosophila staufen protein in the chromosome 8q13-q21.1 region. *Genomics,* 62**,** 113-8.

BYCROFT, M., GRUNERT, S., MURZIN, A. G., PROCTOR, M. & ST JOHNSTON, D. 1995. NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J,* 14**,** 3563-71.

CAMPBELL, I. D. 2002. Timeline: the march of structural biology. *Nature Reviews Molecular Cell Biologyvolume,* 3**,** 377–381.

CANNONE, G., VISENTIN, S., PALUD, A., HENNEKE, G. & SPAGNOLO, L. 2017. Structure of an octameric form of the minichromosome maintenance protein from the archaeon Pyrococcus abyssi. *Sci Rep,* 7**,** 42019.

CASTELLO, A., FISCHER, B., EICHELBAUM, K., HOROS, R., BECKMANN, B. M., STREIN, C., DAVEY, N. E., HUMPHREYS, D. T., PREISS, T., STEINMETZ, L. M., KRIJGSVELD, J. & HENTZE, M. W. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell,* 149**,** 1393-406.

CHATEL-CHAIX, L., BOULAY, K., MOULAND, A. J. & DESGROSEILLERS, L. 2008. The host protein Staufen1 interacts with the Pr55Gag zinc fingers and regulates HIV-1 assembly via its N-terminus. *Retrovirology,* 5**,** 41.

CHATEL-CHAIX, L., CLEMENT, J. F., MARTEL, C., BERIAULT, V., GATIGNOL, A., DESGROSEILLERS, L. & MOULAND, A. J. 2004. Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles. *Mol Cell Biol,* 24**,** 2637-48.

CHEN, Y. J., YU, X., KASIVISWANATHAN, R., SHIN, J. H., KELMAN, Z. & EGELMAN, E. H. 2005. Structural polymorphism of Methanothermobacter thermautotrophicus MCM. *J Mol Biol,* 346**,** 389-94.

CHO, H., HAN, S., PARK, O. H. & KIM, Y. K. 2013. SMG1 regulates adipogenesis via targeting of staufen1-mediated mRNA decay. *Biochim Biophys Acta,* 1829**,** 1276-87.

CHO, H., KIM, K. M., HAN, S., CHOE, J., PARK, S. G., CHOI, S. S. & KIM, Y. K. 2012. Staufen1-mediated mRNA decay functions in adipogenesis. *Mol Cell,* 46**,** 495-506.

COSTA, A., ILVES, I., TAMBERG, N., PETOJEVIC, T., NOGALES, E., BOTCHAN, M. R. & BERGER, J. M. 2011. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol,* 18**,** 471-7.

COSTA, A. & ONESTI, S. 2008. The MCM complex: (just) a replicative helicase? *Biochem Soc Trans,* 36**,** 136-40.

COSTA, A., PAPE, T., VAN HEEL, M., BRICK, P., PATWARDHAN, A. & ONESTI, S. 2006a. Structural basis of the Methanothermobacter thermautotrophicus MCM helicase activity. *Nucleic Acids Res,* 34**,** 5829-38.

COSTA, A., PAPE, T., VAN HEEL, M., BRICK, P., PATWARDHAN, A. & ONESTI, S. 2006b. Structural studies of the archaeal MCM complex in different functional states. *J Struct Biol,* 156**,** 210-9.

COSTA, A., VAN DUINEN, G., MEDAGLI, B., CHONG, J., SAKAKIBARA, N., KELMAN, Z., NAIR, S. K., PATWARDHAN, A. & ONESTI, S. 2008. Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase. *EMBO J,* 27**,** 2250-8.

CRAVEUR, P., JOSEPH, A. P., ESQUE, J., NARWANI, T. J., NOEL, F., SHINADA, N., GOGUET, M., LEONARD, S., POULAIN, P., BERTRAND, O., FAURE, G., REBEHMED, J., GHOZLANE, A., SWAPNA, L. S., BHASKARA, R. M., BARNOUD, J., TELETCHEA, S., JALLU, V., CERNY, J., SCHNEIDER, B., ETCHEBEST, C., SRINIVASAN, N., GELLY, J. C. & DE BREVERN, A. G. 2015. Protein flexibility in the light of structural alphabets. *Front Mol Biosci,* 2**,** 20.

CZISCH, M., BOELENS, R. 1998. Sensitivity Enhancement in the TROSY Experiment. *J. Magn. Reson*. 134, 158-160.

DE LUCAS, S., OLIVEROS, J. C., CHAGOYEN, M. & ORTIN, J. 2014. Functional signature for the recognition of specific target mRNAs by human Staufen1 protein. *Nucleic Acids Res,* 42**,** 4516-26.

DE LUCAS, S., PEREDO, J., MARION, R. M., SANCHEZ, C. & ORTIN, J. 2010. Human Staufen1 protein interacts with influenza virus ribonucleoproteins and is required for efficient virus multiplication. *J Virol,* 84**,** 7603-12.

DEGIACOMI, M. T., IACOVACHE, I., PERNOT, L., CHAMI, M., KUDRYASHEV, M., STAHLBERG, H., VAN DER GOOT, F. G. & DAL PERARO, M. 2013. Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nat Chem Biol,* 9**,** 623-9.

DIXIT, U., PANDEY, A. K., MISHRA, P., SENGUPTA, A. & PANDEY, V. N. 2016. Staufen1 promotes HCV replication by inhibiting protein kinase R and transporting viral RNA to the site of translation and replication in the cells. *Nucleic Acids Res,* 44**,** 5271-87.

DOUCLEFF, M. & CLORE, G. M. 2008. Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc Natl Acad Sci U S A,* 105**,** 13871-6.

DUCHAINE, T., WANG, H. J., LUO, M., STEINBERG, S. V., NABI, I. R. & DESGROSEILLERS, L. 2000. A novel murine Staufen isoform modulates the RNA content of Staufen complexes. *Mol Cell Biol,* 20**,** 5592-601.

DUCHAINE, T. F., HEMRAJ, I., FURIC, L., DEITINGHOFF, A., KIEBLER, M. A. & DESGROSEILLERS, L. 2002. Staufen2 isoforms localize to the somatodendritic domain of neurons and interact with different organelles. *J Cell Sci,* 115**,** 3285-95.

DUGRE-BRISSON, S., ELVIRA, G., BOULAY, K., CHATEL-CHAIX, L., MOULAND, A. J. & DESGROSEILLERS, L. 2005. Interaction of Staufen1 with the 5' end of mRNA facilitates translation of these RNAs. *Nucleic Acids Res,* 33**,** 4797-812.

DUNKER, A. K. & OBRADOVIC, Z. 2001. The protein trinity--linking function and disorder. *Nat Biotechnol,* 19**,** 805-6.

DUNKER, A. K., OBRADOVIC, Z., ROMERO, P., GARNER, E. C. & BROWN, C. J. 2000. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform,* 11**,** 161-71.

DYSON, H. J. & WRIGHT, P. E. 2002. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv Protein Chem,* 62**,** 311-40.

ELBARBARY, R. A., LI, W., TIAN, B. & MAQUAT, L. E. 2013. STAU1 binding 3' UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev,* 27**,** 1495-510.

ERZBERGER, J. P. & BERGER, J. M. 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct,* 35**,** 93-114.

FERNANDEZ MOYA, S. M. & KIEBLER, M. A. 2015. CLIPing Staufen to secondary RNA structures: size and location matter! *Bioessays,* 37**,** 1062-6.

FERNANDEZ-CID, A., RIERA, A., TOGNETTI, S., HERRERA, M. C., SAMEL, S., EVRIN, C., WINKLER, C., GARDENAL, E., UHLE, S. & SPECK, C. 2013. An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Mol Cell,* 50**,** 577-88.

FERRANDON, D., ELPHICK, L., NIISSLEIN-VOLHARD, C. & ST JOHNSTON, D. 1994. Staufen Protein Associates with the 3'UTR of bicoid mRNA to Form Particles That Move in a Microtubule-Dependent Manner
Cell, 79**,** 1221-1232, .

FERRANDON, D., KOCH, I., WESTHOF, E. & NUSSLEIN-VOLHARD, C. 1997. RNA–RNA interaction is required for the formation of specific bicoid mRNA 3 UTR–STAUFEN ribonucleoprotein particles. *The EMBO Journal,* 16**,** 1751–1758.

FLETCHER, R. J., BISHOP, B. E., LEON, R. P., SCLAFANI, R. A., OGATA, C. M. & CHEN, X. S. 2003. The structure and function of MCM from archaeal M. Thermoautotrophicum. *Nat Struct Biol,* 10**,** 160-7.

FLETCHER, R. J. & CHEN, X. S. 2006. Biochemical activities of the BOB1 mutant in Methanobacterium thermoautotrophicum MCM. *Biochemistry,* 45**,** 462-7.

FRIGOLA, J., HE, J., KINKELIN, K., PYE, V. E., RENAULT, L., DOUGLAS, M. E., REMUS, D., CHEREPANOV, P., COSTA, A. & DIFFLEY, J. F. X. 2017. Cdt1 stabilizes an open MCM ring for helicase loading. *Nat Commun,* 8**,** 15720.

FRIGOLA, J., REMUS, D., MEHANNA, A. & DIFFLEY, J. F. 2013. ATPase-dependent quality control of DNA replication origin licensing. *Nature,* 495**,** 339-43.

FURIC, L., MAHER-LAPORTE, M. & DESGROSEILLERS, L. 2008. A genome-wide approach identifies distinct but overlapping subsets of cellular mRNAs associated with Staufen1- and Staufen2-containing ribonucleoprotein complexes. *RNA,* 14**,** 324-35.

GAUTREY, H., MCCONNELL, J., HALL, J. & HESKETH, J. 2005. Polarised distribution of the RNA-binding protein Staufen in differentiated intestinal epithelial cells. *FEBS Lett,* 579**,** 2226-30.

GAUTREY, H., MCCONNELL, J., LAKO, M., HALL, J. & HESKETH, J. 2008. Staufen1 is expressed in preimplantation mouse embryos and is required

for embryonic stem cell differentiation. *Biochim Biophys Acta,* 1783**,** 1935-42.

GEBAUER, F., PREISS, T. & HENTZE, M. W. 2012. From cis-regulatory elements to complex RNPs and back. *Cold Spring Harb Perspect Biol,* 4**,** a012245.

GLEGHORN, M. L., GONG, C., KIELKOPF, C. L. & MAQUAT, L. E. 2013. Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat Struct Mol Biol,* 20**,** 515-24.

GLEGHORN, M. L. & MAQUAT, L. E. 2014. 'Black sheep' that don't leave the double-stranded RNA-binding domain fold. *Trends Biochem Sci,* 39**,** 328-40.

GOKHALE, R. S. & KHOSLA, C. 2000. Role of linkers in communication between protein modules. *Curr Opin Chem Biol,* 4**,** 22-7.

GOMEZ-LLORENTE, Y., FLETCHER, R. J., CHEN, X. S., CARAZO, J. M. & SAN MARTIN, C. 2005. Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from Methanobacterium thermoautotrophicum. *J Biol Chem,* 280**,** 40909-15.

GONG, C., KIM, Y. K., WOELLER, C. F., TANG, Y. & MAQUAT, L. E. 2009. SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. *Genes Dev,* 23**,** 54-66.

GONG, C. & MAQUAT, L. E. 2011. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature,* 470**,** 284-8.

GONG, C., TANG, Y. & MAQUAT, L. E. 2013. mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat Struct Mol Biol,* 20**,** 1214-20.

GOODSELL, D. S. & OLSON, A. J. 2000. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct,* 29**,** 105-53.

HEGYI, H., SCHAD, E. & TOMPA, P. 2007. Structural disorder promotes assembly of protein complexes. *BMC Struct Biol,* 7**,** 65.

HENTZE, M. W., CASTELLO, A., SCHWARZL, T. & PREISS, T. 2018. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol,* 19**,** 327-341.

HERAUD-FARLOW, J. E., SHARANGDHAR, T., LI, X., PFEIFER, P., TAUBER, S., OROZCO, D., HORMANN, A., THOMAS, S., BAKOSOVA, A., FARLOW, A. R., EDBAUER, D., LIPSHITZ, H. D., MORRIS, Q. D.,

BILBAN, M., DOYLE, M. & KIEBLER, M. A. 2013. Staufen2 regulates neuronal target RNAs. *Cell Rep,* 5**,** 1511-8.

HUBBI, M. E., LUO, W., BAEK, J. H. & SEMENZA, G. L. 2011. MCM proteins are negative regulators of hypoxia-inducible factor 1. *Mol Cell,* 42**,** 700-12.

ILVES, I., PETOJEVIC, T., PESAVENTO, J. J. & BOTCHAN, M. R. 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell,* 37**,** 247-58.

IRION, U., ADAMS, J., CHANG, C. W. & ST JOHNSTON, D. 2006. Miranda couples oskar mRNA/Staufen complexes to the bicoid mRNA localization pathway. *Dev Biol,* 297**,** 522-33.

JERABEK-WILLEMSEN, M., ANDRÉ, T., WANNER, R., ROTH, H. M., DUHR, S., BAASKE, P. & BREITSPRECHER, D. 2014. MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure,* 1077**,** 101-113.

JIA, M., SHAN, Z., YANG, Y., LIU, C., LI, J., LUO, Z. G., ZHANG, M., CAI, Y., WEN, W. & WANG, W. 2015. The structural basis of Miranda-mediated Staufen localization during Drosophila neuroblast asymmetric division. *Nat Commun,* 6**,** 8381.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc,* 10**,** 845-58.

KIM, M. Y., PARK, J., LEE, J. J., HA, D. H., KIM, J., KIM, C. G., HWANG, J. & KIM, C. G. 2014. Staufen1-mediated mRNA decay induces Requiem mRNA decay through binding of Staufen1 to the Requiem 3'UTR. *Nucleic Acids Res,* 42**,** 6999-7011.

KIM, Y. K., FURIC, L., DESGROSEILLERS, L. & MAQUAT, L. E. 2005. Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell,* 120**,** 195-208.

KIM, Y. K., FURIC, L., PARISIEN, M., MAJOR, F., DESGROSEILLERS, L. & MAQUAT, L. E. 2007. Staufen1 regulates diverse classes of mammalian transcripts. *The EMBO Journal,* 26**,** 2670–2681.

KRETZ, M. 2013. TINCR, staufen1, and cellular differentiation. *RNA Biol,* 10**,** 1597-601.

KROVAT, B. C. & JANTSCH, M. F. Comparative mutational analysis of the double-stranded RNA binding domains of Xenopus laevis RNA-binding protein A.

LAVER, J. D., LI, X., ANCEVICIUS, K., WESTWOOD, J. T., SMIBERT, C. A., MORRIS, Q. D. & LIPSHITZ, H. D. 2013. Genome-wide analysis of Staufen-associated mRNAs identifies secondary structures that confer target specificity. *Nucleic Acids Res,* 41**,** 9438-60.

LAZZARETTI, D., BANDHOLZ-CAJAMARCA, L., EMMERICH, C., SCHAAF, K., BASQUIN, C., IRION, U. & BONO, F. 2018. The crystal structure of Staufen1 in complex with a physiological RNA sheds light on substrate selectivity. *Life Sci Alliance,* 1**,** e201800187.

LEBEAU, G., MAHER-LAPORTE, M., TOPOLNIK, L., LAURENT, C. E., SOSSIN, W., DESGROSEILLERS, L. & LACAILLE, J. C. 2008. Staufen1 regulation of protein synthesis-dependent long-term potentiation and synaptic function in hippocampal pyramidal cells. *Mol Cell Biol,* 28**,** 2896-907.

LEBOWITZ, J., LEWIS, M. S. & SCHUCK, P. 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci,* 11**,** 2067-79.

LEGENDRE, J. B., CAMPBELL, Z. T., KROLL-CONNER, P., ANDERSON, P., KIMBLE, J. & WICKENS, M. 2013. RNA targets and specificity of Staufen, a double-stranded RNA-binding protein in Caenorhabditis elegans. *J Biol Chem,* 288**,** 2532-45.

LI, P., YANG X FAU - WASSER, M., WASSER M FAU - CAI, Y., CAI Y FAU - CHIA, W. & CHIA, W. Inscuteable and Staufen mediate asymmetric localization and segregation of prospero RNA during Drosophila neuroblast cell divisions.

LIU, J., FAEDER, J. R. & CAMACHO, C. J. 2009. Toward a quantitative theory of intrinsically disordered proteins and their function. *Proc Natl Acad Sci U S A,* 106**,** 19819-23.

LIU, W., PUCCI, B., ROSSI, M., PISANI, F. M. & LADENSTEIN, R. 2008. Structural analysis of the Sulfolobus solfataricus MCM protein N-terminal domain. *Nucleic Acids Res,* 36**,** 3235-43.

LUNDE, B. M., MOORE, C. & VARANI, G. 2007. RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology,* 8**,** 479–490.

LUO, M., DUCHAINE, T. F. & DESGROSEILLERS, L. 2002. Molecular mapping of the determinants involved in human Staufen-ribosome association. *Biochem J,* 365**,** 817-24.

LYUBIMOV, A. Y., COSTA, A., BLEICHERT, F., BOTCHAN, M. R. & BERGER, J. M. 2012. ATP-dependent conformational dynamics underlie the functional asymmetry of the replicative helicase from a minimalist eukaryote. *Proc Natl Acad Sci U S A,* 109**,** 11999-2004.

MACRAE, I. J., LI F FAU - ZHOU, K., ZHOU K FAU - CANDE, W. Z., CANDE WZ FAU - DOUDNA, J. A. & DOUDNA, J. A. 2006. Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harb Symp Quant Biol.*

MAINE, G. T., SINHA, P. & TYE, B. K. 1984. Mutants of S. cerevisiae defective in the maintenance of minichromosomes. *Genetics,* 106**,** 365-85.

MALLARDO, M., DEITINGHOFF, A., MULLER, J., GOETZE, B., MACCHI, P., PETERS, C. & KIEBLER, M. A. 2003. Isolation and characterization of Staufen-containing ribonucleoprotein particles from rat brain. *PNAS,* 100**,** 2100-2105.

MARION, R. M., FORTES, P., BELOSO, A., DOTTI, C. & ORTIN, J. 1999. A human sequence homologue of Staufen is an RNA-binding protein that is associated with polysomes and localizes to the rough endoplasmic reticulum. *Mol Cell Biol,* 19**,** 2212-9.

MARSH, J. A. & TEICHMANN, S. A. 2015. Structure, dynamics, assembly, and evolution of protein complexes. *Annu Rev Biochem,* 84**,** 551-75.

MARTEL, C., DUGRE-BRISSON, S., BOULAY, K., BRETON, B., LAPOINTE, G., ARMANDO, S., TREPANIER, V., DUCHAINE, T., BOUVIER, M. & DESGROSEILLERS, L. 2010. Multimerization of Staufen1 in live cells. *RNA,* 16**,** 585-97.

MARTEL, C., MACCHI, P., FURIC, L., KIEBLER, M. A. & DESGROSEILLERS, L. 2006. Staufen1 is imported into the nucleolus via a bipartite nuclear localization signal and several modulatory determinants. *Biochem J,* 393**,** 245-54.

MEAGHER, M., EPLING, L. B. & ENEMARK, E. J. 2019. DNA translocation mechanism of the MCM complex and implications for replication initiation. *Nat Commun,* 10**,** 3117.

MICKLEM, D. R., ADAMS, J., GRUNERT, S. & ST JOHNSTON, D. 2000. Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. *The EMBO Journal,* 19**,** 1366–1377.

MILLER, J. M. & ENEMARK, E. J. 2015. Archaeal MCM Proteins as an Analog for the Eukaryotic Mcm2-7 Helicase to Reveal Essential Features of Structure and Function. *Archaea,* 2015**,** 305497.

MOULAND, A. J., MERCIER, J., LUO, M., BERNIER, L., DESGROSEILLERS, L. & COHEN, E. A. 2000. The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation. *J Virol,* 74**,** 5441-51.

NOGUCHI, Y., YUAN, Z., BAI, L., SCHNEIDER, S., ZHAO, G., STILLMAN, B., SPECK, C. & LI, H. 2017. Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-strand DNA extrusion model. *Proc Natl Acad Sci U S A,* 114**,** E9529-E9538.

OGINO, H., ISHINO, S., HAUGLAND, G. T., BIRKELAND, N. K., KOHDA, D. & ISHINO, Y. 2014. Activation of the MCM helicase from the thermophilic archaeon, Thermoplasma acidophilum by interactions with GINS and Cdc6-2. *Extremophiles,* 18**,** 915-24.

OTWINOWSKI, Z., SCHEVITZ, R. W., ZHANG, R. G., LAWSON, C. L., JOACHIMIAK, A., MARMORSTEIN, R. Q., LUISI, B. F. & SIGLER, P. B. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature,* 335**,** 321-9.

PABON, N. A. & CAMACHO, C. J. 2017. Probing protein flexibility reveals a mechanism for selective promiscuity. *Elife,* 6.

PAN, M., SANTANGELO, T. J., LI, Z., REEVE, J. N. & KELMAN, Z. 2011. Thermococcus kodakarensis encodes three MCM homologs but only one is essential. *Nucleic Acids Res,* 39**,** 9671-80.

PAPALEO, E., SALADINO, G., LAMBRUGHI, M., LINDORFF-LARSEN, K., GERVASIO, F. L. & NUSSINOV, R. 2016. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem Rev,* 116**,** 6391-423.

PARK, E., GLEGHORN, M. L. & MAQUAT, L. E. 2013. Staufen2 functions in Staufen1-mediated mRNA decay by binding to itself and its paralog and promoting UPF1 helicase but not ATPase activity. *Proc Natl Acad Sci U S A,* 110**,** 405-12.

PARK, E. & MAQUAT, L. E. 2013. Staufen-mediated mRNA decay. *Wiley Interdiscip Rev RNA,* 4**,** 423-35.

PARK, E. G., M.L.; MAQUAT, L.E. 2013. Staufen2 functions in Staufen1-mediated mRNA decay by binding to itself and its paralog and promoting UPF1 helicase but not ATPase activity. *PNAS,* 110**,** 405–412.

PEREDO, J., VILLACE, P., ORTIN, J. & DE LUCAS, S. 2014. Human Staufen1 associates to miRNAs involved in neuronal cell differentiation and is required for correct dendritic formation. *PLoS One,* 9**,** e113704.

PERICA, T., CHOTHIA, C. & TEICHMANN, S. A. 2012. Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc Natl Acad Sci U S A,* 109**,** 8127-32.

PERVUSHIN, K., WIDER, G., WUETHRICH, K. 1998. Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *J. Biomol. NMR,* 12, 345-348.

PHAN, A. T., KURYAVYI, V., DARNELL, J. C., SERGANOV, A., MAJUMDAR, A., ILIN, S., RASLIN, T., POLONSKAIA, A., CHEN, C., CLAIN, D., DARNELL, R. B. & PATEL, D. J. 2011. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol,* 18**,** 796-804.

QUADE, N., BOEHRINGER, D., LEIBUNDGUT, M., VAN DEN HEUVEL, J. & BAN, N. 2015. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-A resolution. *Nat Commun,* 6**,** 7646.

RALSTON, G. L. 2004. Introduction to analytical ultracentrifugation. *Conference Proceedings*.

RAMASAMY, S., WANG, H., QUACH, H. N. & SAMPATH, K. 2006. Zebrafish Staufen1 and Staufen2 are required for the survival and migration of primordial germ cells. *Dev Biol,* 292**,** 393-406.

RAMBO, R. P. 2017. Considerations for Sample Preparation Using Size-Exclusion Chromatography for Home and Synchrotron Sources. *Adv Exp Med Biol.*

RAMOS, A., BAYER, P. & VARANI, G. 2000a. Determination of the Structure of the RNA Complex of a Double- Stranded RNA-Binding Domain from Drosophila Staufen Protein. *Biopolymers (Nucleic Acid Sciences),* 52**,** 181-196.

RAMOS, A., GRU¨NERT, S., ADAMS, J., MICKLEM, D. R., PROCTOR, M. R., FREUND, S., BYCROFT, M., ST JOHNSTON, D. & VARANI, G. 2000b. RNA recognition by a Staufen double-stranded RNA-binding domain. *The EMBO Journal,* 19**,** 997–1009.

RAMOS, A., GRUNERT, S., ADAMS, J., MICKLEM, D. R., PROCTOR, M. R., FREUND, S., BYCROFT, M., ST JOHNSTON, D. & VARANI, G. 2000c. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J,* 19**,** 997-1009.

RAVEL-CHAPUIS, A., BELANGER, G., YADAVA, R. S., MAHADEVAN, M. S., DESGROSEILLERS, L., COTE, J. & JASMIN, B. J. 2012. The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing. *J Cell Biol,* 196**,** 699-712.

RAVEL-CHAPUIS, A., CRAWFORD, T. E., BLAIS-CREPEAU, M. L., BELANGER, G., RICHER, C. T. & JASMIN, B. J. 2014. The RNA-binding protein Staufen1 impairs myogenic differentiation via a c-myc-dependent mechanism. *Mol Biol Cell,* 25**,** 3765-78.

RAVEL-CHAPUIS, A., KLEIN GUNNEWIEK, A., BELANGER, G., CRAWFORD PARKS, T. E., COTE, J. & JASMIN, B. J. 2016. Staufen1 impairs stress granule formation in skeletal muscle cells from myotonic dystrophy type 1 patients. *Mol Biol Cell,* 27**,** 1728-39.

REMUS, D., BEURON, F., TOLUN, G., GRIFFITH, J. D., MORRIS, E. P. & DIFFLEY, J. F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell,* 139**,** 719-30.

RICCI, E. P., KUCUKURAL, A., CENIK, C., MERCIER, B. C., SINGH, G., HEYER, E. E., ASHAR-PATEL, A., PENG, L. & MOORE, M. J. 2014. Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol,* 21**,** 26–35.

SAMSON, R. Y., ABEYRATHNE, P. D. & BELL, S. D. 2016. Mechanism of Archaeal MCM Helicase Recruitment to DNA Replication Origins. *Mol Cell,* 61**,** 287-96.

SAMSON, R. Y. & BELL, S. D. 2016. Archaeal DNA Replication Origins and Recruitment of the MCM Replicative Helicase. *Enzymes,* 39**,** 169-90.

SLATTERY, M., RILEY, T., LIU, P., ABE, N., GOMEZ-ALCALA, P., DROR, I., ZHOU, T., ROHS, R., HONIG, B., BUSSEMAKER, H. J. & MANN, R. S. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell,* 147**,** 1270-82.

SLAYMAKER, I. M., FU, Y., TOSO, D. B., RANATUNGA, N., BREWSTER, A., FORSBURG, S. L., ZHOU, Z. H. & CHEN, X. S. 2013. Mini-chromosome maintenance complexes form a filament to remodel DNA structure and topology. *Nucleic Acids Res,* 41**,** 3446-56.

ST JOHNSTON, D., BEUCHLE, D. & NIISSLEIN-VOLHARD, C. 1991. staufen, a Gene Required to Localize Maternal RNAs in the Drosophila Egg. *Cell,* 6**,** 51-63.

ST JOHNSTON, D., BROWN, N. H., GALL, J. G. & JANTSCH, M. 1992. A conserved double-stranded RNA-binding domain. *Proc. Natl. Acad. Sci. USA,* 89**,** 10979-10983.

ST JOHNSTON, D. & NUSSLEIN-VOLHARD, C. 1992. The origin of pattern and polarity in the Drosophila embryo. *Cell,* 68**,** 201-19.

STOCK, D., LESLIE, A. G. & WALKER, J. E. 1999. Molecular architecture of the rotary motor in ATP synthase. *Science,* 286**,** 1700-5.

SUGIMOTO, Y., VIGILANTE, A., DARBO, E., ZIRRA, A., MILITTI, C., D'AMBROGIO, A., LUSCOMBE, N. M. & ULE, J. 2015. hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature,* 519**,** 491-4.

SVERGUN, D. I., KOCH, M. H. J., TIMMINS, P. A. & MAY, R. P. 2013. *Small Angle X-ray and Neutron Scattering from Solutions of Biological Macromolecules* Oxford University Press.

TAKAHASHI, T. S., WIGLEY, D. B. & WALTER, J. C. 2005. Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem Sci,* 30**,** 437-44.

THOMAS, J. M. & BEAL, P. A. 2017. How do ADARs bind RNA? New protein-RNA structures illuminate substrate recognition by the RNA editing ADARs. *BioEssays : news and reviews in molecular, cellular and developmental biology,* 39**,** 10.1002/bies.201600187.

TOMPA, P. 2011. Unstructural biology coming of age. *Curr Opin Struct Biol,* 21**,** 419-25.

TOMPA, P. & FUXREITER, M. 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci,* 33**,** 2-8.

TOMPA, P., SCHAD, E., TANTOS, A. & KALMAR, L. 2015. Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol,* 35**,** 49-59.

TRIA, G., MERTENS, H. D. T., KACHALA, M. & SVERGUN, D. I. 2015. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ,* 2**,** 207-217.

TRNKA, M. J., PELLARIN, R. & ROBINSON, P. J. 2019. Role of integrative structural biology in understanding transcriptional initiation. *Methods,* 159-160**,** 4-22.

TSVETANOVA, N. G., KLASS, D. M., SALZMAN, J. & BROWN, P. O. 2010. Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. *PLoS One,* 5.

TYE, B. K. 1999. MCM proteins in DNA replication. *Annu Rev Biochem,* 68**,** 649-86.

VARADI, M., ZSOLYOMI, F., GUHAROY, M. & TOMPA, P. 2015. Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. *PLoS One,* 10**,** e0139731.

VASILYEV, N., POLONSKAIA, A., DARNELL, J. C., DARNELL, R. B., PATEL, D. J. & SERGANOV, A. 2015. Crystal structure reveals specific recognition of a G-quadruplex RNA by a beta-turn in the RGG motif of FMRP. *Proc Natl Acad Sci U S A,* 112**,** E5391-400.

VESSEY, J. P., MACCHI, P., STEIN, J. M., MIKL, M., HAWKER, K. N., VOGELSANG, P., WIECZOREK, K., VENDRA, G., RIEFLER, J., TU¨ BING, F., APARICIO, S. A. J., ABEL, T. & KIEBLER, M. A. 2008. A loss of function allele for murine Staufen1 leads to impairment of dendritic Staufen1-RNP delivery and dendritic spine morphogenesis. *PNAS,* 105**,** 16374-16379.

VOGEL, C., BASHTON, M., KERRISON, N. D., CHOTHIA, C. & TEICHMANN, S. A. 2004. Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol,* 14**,** 208-16.

VRANKEN, W. F., BOUCHER, W., STEVENS, T. J., FOGH, R. H., PAJON, A., LLINAS, M., ULRICH, E. L., MARKLEY, J. L., IONIDES, J. & LAUE, E. D. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins,* 59**,** 687-96.

WANG, X., VUKOVIC, L., KOH, H. R., SCHULTEN, K. & MYONG, S. 2015. Dynamic profiling of double-stranded RNA binding proteins. *Nucleic Acids Res,* 43**,** 7566-76.

WICKHAM, L., DUCHAINE, T., LUO, M., NABI, I. R. & DESGROSEILLERS, L. 1999a. Mammalian Staufen Is a Double-Stranded-RNA- and Tubulin-Binding Protein Which Localizes to the Rough Endoplasmic Reticulum. *Molecular and Cellular Biology,* 19**,** 2220-2230.

WICKHAM, L., DUCHAINE, T., LUO, M., NABI, I. R. & DESGROSEILLERS, L. 1999b. Mammalian staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum. *Mol Cell Biol,* 19**,** 2220-30.

WRIGHT, P. E. & DYSON, H. J. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol,* 16**,** 18-29.

XIA, Y., NIU, Y., CUI, J., FU, Y., CHEN, X. S., LOU, H. & CAO, Q. 2015. The Helicase Activity of Hyperthermophilic Archaeal MCM is Enhanced at High Temperatures by Lysine Methylation. *Front Microbiol,* 6**,** 1247.

YOSHIMOCHI, T., FUJIKANE, R., KAWANAMI, M., MATSUNAGA, F. & ISHINO, Y. 2008. The GINS complex from Pyrococcus furiosus stimulates the MCM helicase activity. *J Biol Chem,* 283**,** 1601-9.

YU, L., REUTZEL-EDENS, S. M. & MITCHELL, C. A. 2000. Crystallization and Polymorphism of Conformationally Flexible Molecules: Problems, Patterns, and Strategies. *Organic Process Research & Development,* 4**,** 396-402.

ZOU, L., MITCHELL, J. & STILLMAN, B. 1997. CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. *Mol Cell Biol,* 17**,** 553-63.

Appendix

List of Publications

ACS NANO
Cite This: ACS Nano 2018, 12, 5834–5847
www.acsnano.org

ARTICLE

C1q-Mediated Complement Activation and C3 Opsonization Trigger Recognition of Stealth Poly(2-methyl-2-oxazoline)-Coated Silica Nanoparticles by Human Phagocytes

Regina Tavano,[†] Luca Gabrielli,[‡] Elisa Lubian,[‡] Chiara Fedeli,[†,▽] Silvia Visentin,[†,▼] Patrizia Polverino De Laureto,[§] Giorgio Arrigoni,[†] Alessandra Geffner-Smith,[†] Fangfang Chen,[∥,⊥] Dmitri Simberg,[∥] Giulia Morgese,[#] Edmondo M. Benetti,[#] Linping Wu,[¶,△] Seyed Moein Moghimi,*[,∥,△,▲] Fabrizio Mancin,[‡] and Emanuele Papini*[,†]

[†]Department of Biomedical Sciences, [‡]Department of Chemical Sciences, and [§]Department of Pharmaceutical Sciences, University of Padua, Padua 35121, Italy

[∥]Translational Bio-Nanosciences Laboratory and Colorado Center for Nanomedicine and Nanosafety, The Skaggs School of Pharmacy and Pharmaceutical Sciences, Department of Pharmaceutical Sciences, University of Colorado Denver, Anschutz Medical Campus, 1250 East Mountview Boulevard, Aurora, Colorado 80045, United States

[⊥]Department of Gastrointestinal Surgery, China-Japan Union Hospital, Jilin University, 126 Xiantai Street, Changchun, Jilin 130033, China

[#]Department of Materials, ETH, Zurich CH-8093, Switzerland

[¶]Guangzhou Institute of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

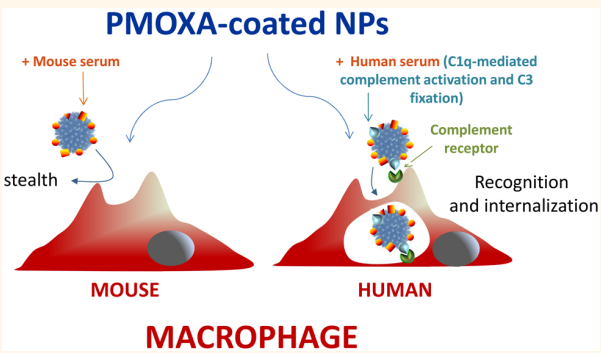
[△]School of Pharmacy, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom

[▲]Institute of Cellular Medicine, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom

Supporting Information

ABSTRACT: Poly(2-methyl-2-oxazoline) (PMOXA) is an alternative promising polymer to poly(ethylene glycol) (PEG) for design and engineering of macrophage-evading nanoparticles (NPs). Although PMOXA-engineered NPs have shown comparable pharmacokinetics and *in vivo* performance to PEGylated stealth NPs in the murine model, its interaction with elements of the human innate immune system has not been studied. From a translational angle, we studied the interaction of fully characterized PMOXA-coated vinyltriethoxysilane-derived organically modified silica NPs (PMOXA-coated NPs) of approximately 100 nm in diameter with human complement system, blood leukocytes, and macrophages and compared their performance with PEGylated and uncoated NP counterparts. Through detailed immunological and proteomic profiling, we show that PMOXA-coated NPs extensively trigger complement activation in human sera exclusively through the classical pathway. Complement activation is initiated by the sensing molecule C1q, where C1q binds with high affinity (K_d = 11 ± 1 nM) to NP surfaces independent of immunoglobulin binding. C1q-mediated complement activation accelerates PMOXA opsonization with the third complement protein (C3) through the amplification loop of the alternative pathway. This promoted NP recognition by human blood leukocytes and monocyte-derived macrophages. The macrophage capture of PMOXA-coated NPs correlates with sera donor variability in complement activation and opsonization but not with other major corona proteins, including clusterin and a wide range of

continued...



Received: March 9, 2018
Accepted: May 11, 2018
Published: May 11, 2018

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834–5847

apolipoproteins. In contrast to these observations, PMOXA-coated NPs poorly activated the murine complement system and were marginally recognized by mouse macrophages. These studies provide important insights into compatibility of engineered NPs with elements of the human innate immune system for translational steps.

KEYWORDS: *C1q, C3, complement, human macrophages, polyoxazoline, stealth polymers*

The interaction of intravenously injected nanoparticles (NPs) with the host blood proteins has long been suggested to modulate NP pharmacokinetics and their extent of clearance from systemic circulation by macrophages of the reticuloendothelial system (RES).[1,2] The rapid clearance of therapeutic NPs by macrophages of the RES, however, could limit their delivery to target sites outside the liver and the spleen.[1−4] Steric stabilization of NPs with long-chain hydrophilic polymers/copolymers has become a successful approach in minimizing protein binding and modulating NP pharmacokinetics and biodistribution. Among many engineered polymers, poly(ethylene glycol) (PEG) has widely been used for design and surface engineering of long-circulating NPs and vesicular drug carriers.[5] A classic example is the regulatory-approved Doxil (a PEGylated liposomal formulation of doxorubicin) and its generic versions.[6]

Although PEGylation can successfully reduce adsorption of blood proteins to NPs, numerous studies have shown that opsonization of PEGylated NPs by the third protein of the complement system (C3) and its cleavage products (C3b/iC3b) may still occur, and this could lead to their efficient capture by human macrophages.[7,8] It is also crucial to consider that some individuals may have antibodies that could recognize the PEG moiety of PEGylated nanopharmaceuticals. This may initiate macrophage recognition of PEGylated NPs through complement opsonization.[9−13] Furthermore, PEG can undergo oxidative degradation, and this may increase its complement activation property.[14] A number of animal and clinical studies have also indicated that PEGylated NPs may induce cardiopulmonary disturbances and distress. Earlier studies postulated that complement activation and subsequent release of C3a and C5a anaphylotoxins induced by PEGylated NPs may play a causal role in initiating such reactions.[7,8,15,16] However, a recent study has demonstrated a transitional link between robust NP clearance from the blood by strategically placed macrophages in vasculature and adverse hemodynamic reactions independent of complement activation.[17] This further indicates that PEGylated NPs may not necessarily behave as stealth entities depending on microenvironmental conditions.

In line with the above-mentioned issues, alternative strategies are being considered for design and engineering of long-circulating and macrophage-evading NPs.[18] Among many approaches in NP surface engineering is surface functionalization with polyoxazoline polymers and their derivatives, which has generated promising outcomes.[19−27] For instance, polyoxazoline-coated NPs show considerable resistance to protein binding and are more amenable to further chemical modifications compared with PEGylated NPs.[28−30] Furthermore, polyoxazoline-coated NPs exhibit prolonged circulation profiles on intravenous injection in the murine model with comparable pharmacokinetic profiles to PEGylated NPs.[27,31,32] Murine models and murine- and bovine-derived materials are widely used in nanomedicine evaluations. However, there are considerable species differences (*e.g.*, mouse *versus* human) in innate immune responses to particulate invaders, including complement activation, C3 opsonization processes, and phagocytic clearance. For instance, although the uptake of superparamagnetic iron oxide nanoworms by both murine and human leukocytes is C3-dependent, there are major differences in pathways of complement activation and the extent of C3 fixation between these species.[33] Accordingly, species differences in innate immune system function and performance can modulate NP pharmacokinetics and responses differently. Therefore, prior to translational and clinical studies, it is necessary to confirm stealth characteristics of engineered nanopharmaceuticals, at least with respect to the human complement system, blood leukocytes, and macrophage responses. Here, we have tested stealth characteristics of poly(2-methyl-2-oxazoline)-coated vinyltriethoxysilane-derived organically modified silica NPs (PMOXA-coated NPs) in human sera from different individuals against complement activation, complement opsonization, and dysopsonization processes and capturing efficacy by human blood leukocytes and monocyte-derived macrophages. Our approach has considered interindividual responses and highlights important insights into the mechanisms of compatibility of nanomaterials with elements of the human innate immunity and disparity with the murine system.

RESULTS AND DISCUSSION

Synthesis and Physicochemical Properties of NPs. Polymeric NPs were prepared by ammonia-catalyzed microemulsion polymerization of vinyltrietoxysilane (VTES) (Figure 1A and Figures S1−S3).[34] Fluorescent labeling and surface functionalization with PEG (M_w = 2000 Da, degree of polymerization = 44) and PMOXA (M_w = 4000 Da, degree of polymerization = 40) were achieved by copolymerizing VTES with Rhodamine B triethoxysilane and, when needed, with the trimethoxysilane derivatives of the polymers. ^{1}H NMR confirmed reaction completion and the absence of byproducts. Broadening of the ^{1}H NMR polymer signals (with respect to their corresponding linewidth in solution), as well as diffusion filter experiments, confirmed NP grafting with designated polymers (Figure S4). All NP types exhibited spheroidal morphology (determined by transmission electron microscopy, Figure S4) with calculated diameter (mean ± SD) of 115 ± 23, 90 ± 10, and 70 ± 6 (n = 400) for uncoated, PEGylated, and PMOXA-coated species, respectively. NP hydrodynamic size distribution and concentration were also measured by nanoparticle tracking analysis (NTA). This modality overcomes intrinsic problems observed when DLS is applied to heterogeneous samples because it is based on video tracking of the Brownian motion of single NPs.[35] The NTA results revealed mean hydrodynamic diameters (mean ± SD) of 144 ± 5 (mode 124 ± 8), 117 ± 4 (mode 103 ± 2), and 86 ± 4 (mode 83 ± 2) for uncoated, PEGylated, and PMOXA-coated species (n = 3 measurements in all cases), respectively. Size distributions were almost symmetrical, with a ±D50% of 20−25 nm for all NPs. Dynamic laser light scattering yielded results comparable to those with NTA for all NP preparations with polydispersity indices <0.04, thus confirming near monodisperse NP suspensions. The small differences found in nanoparticle sizes are likely the result of the different additives (*i.e.*, the alkoxysilane derivatives of the coating polymers) present

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

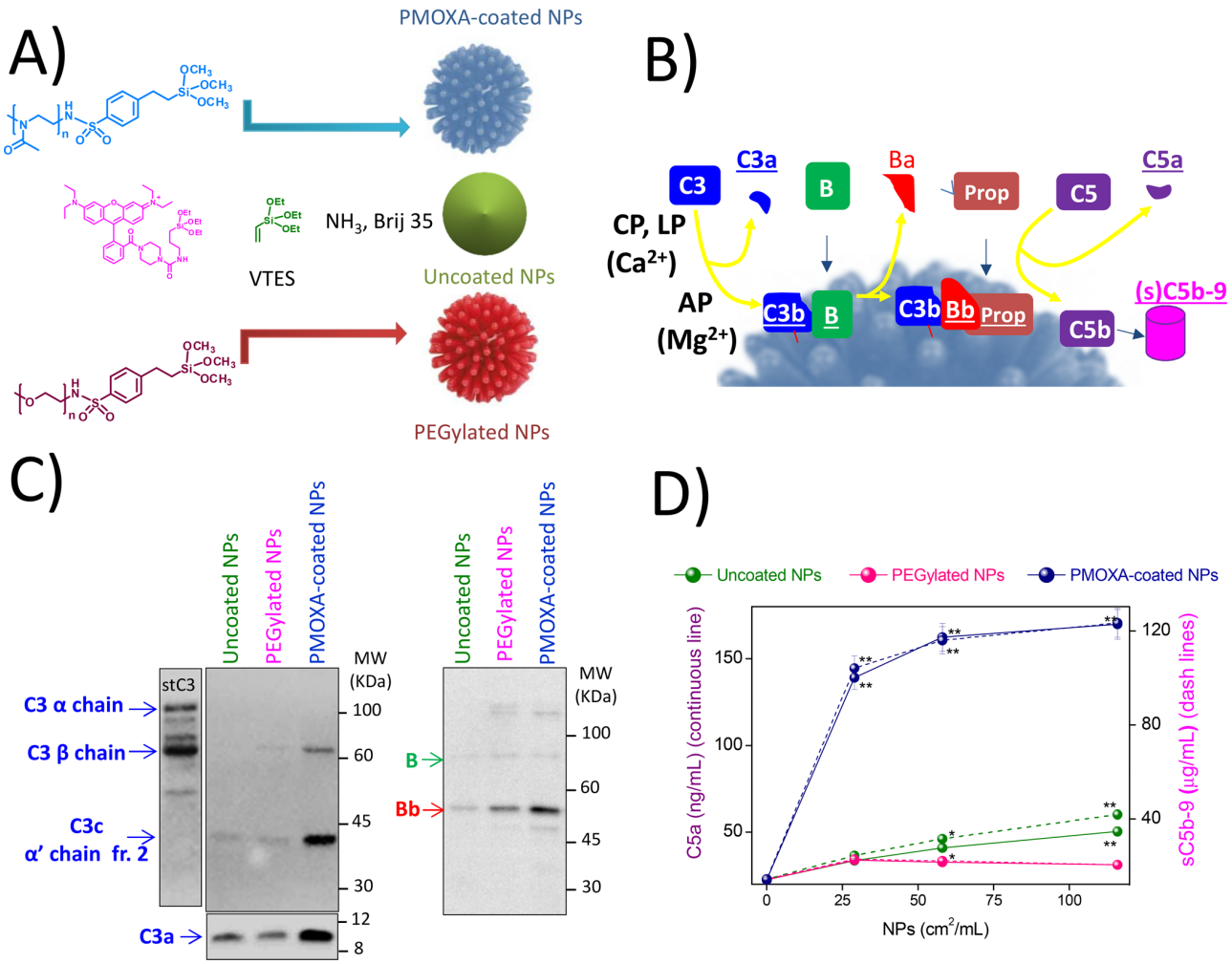


Figure 1. Complement activation by engineered NPs in human serum. (A) Schematic representation of NP structure. (B) Simplified scheme of complement activation pathways (CP, classical pathway; LP, lectin pathway; AP, alternative pathway). The scheme depicts C3 conversion into C3b through assembly of pathway-dependent C3 convertases and generation of measurable complement activation products such as C3a, Bb, C5a, and sC5b-9. (C) Western blot analysis of complement activation products by NPs. The panel shows detection of C3a (M_w ~ 9 kDa) in the fluid phase and surface-bound C3 α (M_w ~ 110 kDa), C3 β (M_w ~ 70 kDa) chains, C3c α' chain fragment 2 (a fragment of C 3d, M_w ~ 39 kDa), factor B (M_w ~ 85 kDa), and Bb (M_w ~ 55 kDa) from washed NPs. StC3 represents purified human C3. (D) Determination of NP-mediated complement activation as a function of NP surface area by ELISA. Complement activation was monitored through measurements of fluid phase C5a and sC5b-9. Data are mean ± SE (n = 3). Statistical analyses were performed with student's t test to calculate significance (*p < 0.05, **p < 0.01) compared with corresponding controls. Zymosan (200 μg/mL) was used as the positive control for monitoring complement activation, resulting in formation of 249 ng/mL serum of C5a and 35707 ng/mL serum of sC5b-9. The maximal effect of PMOXA NPs correspond to 67% (C5a) and 35% (sC5b-9) of the Zymosan activity.

in the reaction mixtures. Thermogravimetric analysis indicated approximately 13 000 PEG molecules and 16 000 PMOXA molecules per designated NP. These values correspond to surface footprints of 2.8 and 2.2 nm^2 for PEG and PMOXA, respectively, and to a distance between grafting points of 1.9 and 1.7 nm, respectively, which indicate that both the NP types feature a dense brush shell and that differences in their interfacial properties are solely due to their diverse shell composition (Table S4). Calculated ζ-potential values of all engineered polymer-coated NPs (−4 mV) were comparable with those of uncoated NPs (−6 mV). All NP types exhibited comparable UV−visible spectra (Figure S4), and Rhodamine B loadings were in the 0.1−0.3% w/w range.

Complement Activation Studies. NP-mediated complement activation in human serum (HS) was followed through measurement of a number of established complement activation products bound to NP surfaces as well as in the aqueous phase (see scheme in Figure 1B). Based on Western blot analysis, PMOXA-coated NPs were more effective in liberating C3a (the anaphylatoxin released on C3 cleavage) than both PEGylated and uncoated NPs (Figure 1C). The potency of PMOXA-coated NPs in activating the human complement system was further confirmed by showing the abundant presence of C3c α' chain fragment 2, which is generated after conversion of C3b into the inactive C3b (iC3b), confirming C3 convertase activity as well as Bb generation (confirming the involvement of the alternative pathway of the complement system) compared with other NP species (Figure 1C). In addition to these, ELISA studies showed that PMOXA-coated NPs are far more efficient in liberating two markers of the terminal pathway of the human complement

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

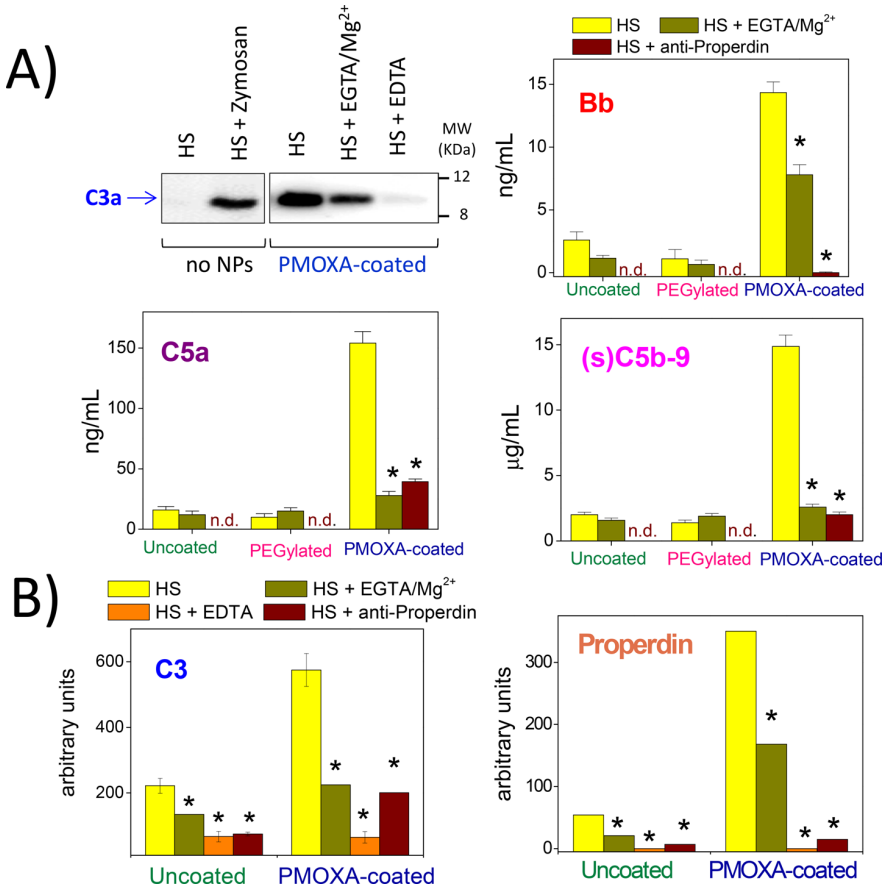


Figure 2. Effect of divalent cation chelation and antiproperdin antibodies on complement activation by engineered NPs in HS. (A) Western blot representation (repeated three times) of C3a generation by PMOXA-coated NPs in normal HS and after serum treatment with chelating agents (top left panel). Control lanes include untreated HS (negative control) and HS incubated with Zymosan (100 μg/mL; positive control). Other panels show NP (119 cm^2/mL)-mediated C5a, Bb, and sC5b-9 release (determined by ELISA) before and after serum treatment with either chelating agents or antiproperdin antibodies. Data are the mean ± SE (n = 3). (B) Densitometric quantification (arbitrary units) of dot blot analysis of C3 and properdin binding to engineered NPs on incubation with untreated, chelated, and antiproperdin antibody-treated HS. Values are mean ± SE (n = 3); *statistical significance (p < 0.05) compared with respective controls.

system (the anaphylatoxin C5a and sC5b-9, which is the soluble form of the membrane attack complex) than PEGylated and uncoated NPs on the basis of equivalent surface area (Figure 1D).

Complement Activation Pathways Triggered by PMOXA-Coated NPs. PMOXA-coated NP-mediated Bb liberation confirms a role for the alternative pathway; however, it is not clear whether complement activation is solely arising from this pathway or if there is a role for both classical and lectin pathways and/or the amplification loop of the alternative pathway. It is well-known that the activation of the classical and lectin pathways of the complement system is Ca^{2+}-dependent, whereas Mg^{2+} is essential for the operation of the alternative pathway.[36,37] We found that selective chelation of serum Ca^{2+} (ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*',*N*'-tetraacetic acid (EGTA)/Mg^{2+} treatment) strongly reduces PMOXA-coated NP-mediated C3a release, whereas chelation of both Ca^{2+} and Mg^{2+} (ethylenediaminetetraacetic acid (EDTA) treatment) abolishes anaphylatoxin release (Figure 2A). Consistent with these observations, PMOXA-coated NP-induced generation of both C5a and sC5b-9 is dramatically reduced on Ca^{2+} chelation. On the other hand, prevention of the alternative pathway convertase stabilization by a neutralizing antiproperdin antibody abolished Bb generation and dramatically reduced C5a release and sC5b-9 formation on PMOXA-coated NP treatment. Furthermore, dot immunoblot studies (Figure 2B) showed that the binding of both C3 and properdin to PMOXA-coated NPs is affected by EGTA/Mg^{2+} and essentially eliminated by EDTA. Collectively, these observations indicate that complement activation by PMOXA-coated NPs is predominantly initiated through Ca^{2+}-sensitive pathways. Subsequently, this increases the turnover of the amplification loop of the alternative pathway to maximize C3 opsonization through assembly of the alternative pathway C3 convertase. In contrast to PMOXA-coated NPs, complement activation by uncoated and PEGylated NPs is marginal.

Proteomic Analysis of Complement Protein Deposition on NPs. Next, we performed shot-gun proteomics after treatment of NPs with normal and chelated HS to gain more insights into complement activation processes (Figure 3A, Figures S5 and S6, and SI files 1−3). Among many deposited species, we detected several effector and regulatory components of the complement system (predominantly C1q, C1r, C1s, C3, C4, C5, C6, C7, C8, C9, B, P, C4Bp, CFAH, and related isoforms) on PMOXA-coated NPs. The surface presence of complement proteins was more prominent on PMOXA-coated NPs than PEGylated and uncoated NPs and in line with functional complement activation studies (Figure S6).

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

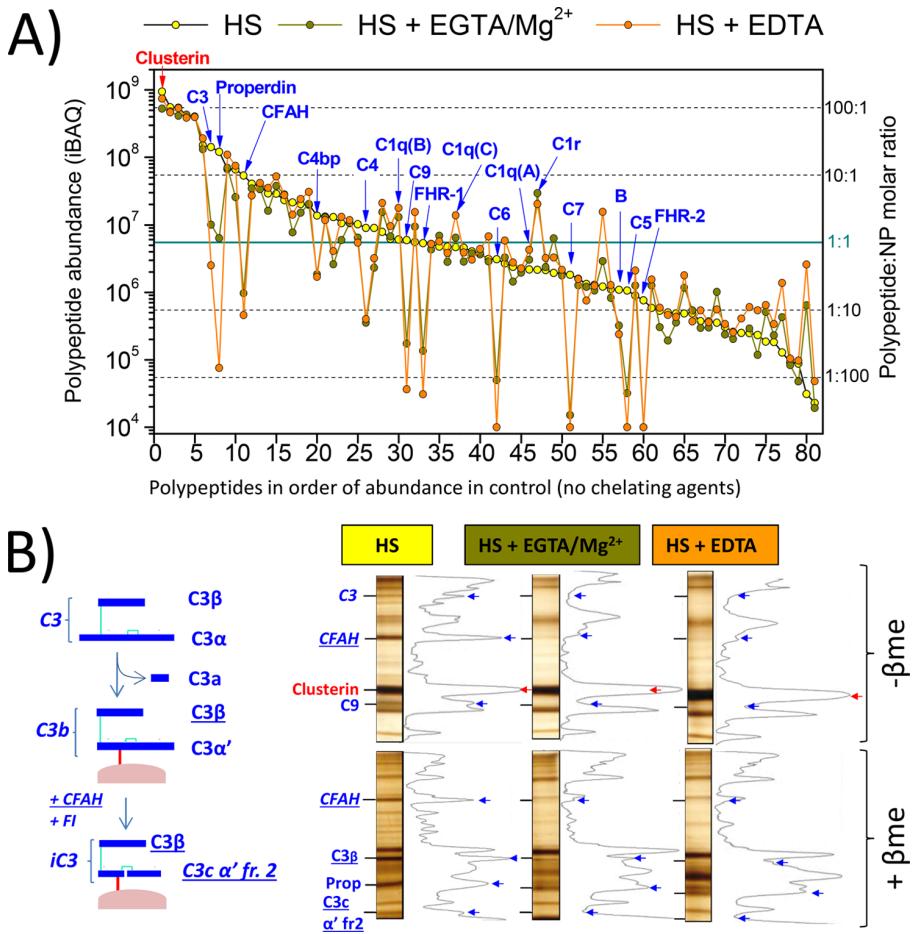


Figure 3. Proteomics analysis of complement proteins on engineered NPs. (A) Shot-gun and label-free quantification of complement proteins from protein corona of PMOXA-coated NPs after incubation in untreated and chelated HS. The iBAQ values (left axis) of identified polypeptides were plotted in order of abundance in control samples. Right axis indicates approximate stoichiometry per NP for each polypeptide. Complement proteins are indicated by blue arrows. Clusterin is shown in red. (B) Mass spectrometry after SDS-PAGE, in-gel digestion, and densitometry. The representative panel shows the identification of major complement proteins and their fragments (in blue) in the corona of PMOXA-coated NPs (exemplified in the pictogram on the left). NP-bound proteins are highlighted by densitometry after nonreducing (−β*me*) and reducing (+β*me*) electrophoresis and silver staining. Peaks corresponding to complement proteins (and their fragments) are indicated by blue arrows. The clusterin band is indicated in red.

However, there was a discrepancy between shot-gun and functional studies. Whereas functional studies confirmed a role for calcium-sensitive pathways, the second complement protein (C2) was not detectable by shot-gun profiling. Nevertheless, the general trend of shot-gun proteomics is in agreement with functional studies involving chelators as binding of the majority of complement proteins is dramatically reduced (>90%) on calcium chelation and further inhibited on EDTA treatment of serum (Figure 3A and Figures S5 and S6). In contrast to complement proteins, other predominantly deposited species such as clusterin, lipoproteins (LPs), serum albumin, and immunoglobulins (Igs), are marginally affected on divalent cation chelation, regardless of NP type (Figure 3A and Figures S5 and S6).

MS spectrometry analysis, after in-gel digestions of major electrophoretic bands obtained by separation of the NP-bound HS polypeptides by SDS-PAGE, further agreed with shot-gun analysis (Figure 3B, Figure S7, and SI files 4−7). Notably, the bands corresponding to C3 β chain ($M_w \sim 65$ kDa) and C3c α' chain fragment 2 ($M_w \sim 39$ kDa) were only detected under reducing conditions, which confirms C3 cleavage and covalent association of C3b with NP surfaces. Densitometry analysis of samples incubated in the presence of chelating agents, once again, showed that only the binding of complement proteins to PMOXA-coated NPs is strongly reduced on Ca^{2+} chelation and further inhibited on EDTA treatment (Figure 3B).

Role of Human Serum and Complement Fixation on PMOXA-Coated NP Uptake by Human Phagocytes. Circulating blood phagocytes (monocytes and PMNGs (polymorphonuclear granulocytes)) and human macrophages showed preferential tendency to capture PMOXA-coated NPs and PEGylated NPs, compared with uncoated NPs, pretreated with HS (75% v/v) for 15 min at 37 °C in the absence of cells and then incubated with cells after a 10-fold dilution in protein-free culture medium (Figure 4A,B). Confocal fluorescent microscopy confirmed internalization of all preopsonized NPs and their localization to the acidic endosomal−lysosomal compartments (Figure 4C). Cell viability was further confirmed on NP uptake studies (Figure S8). Preopsonized NP uptake by both blood phagocytes and human macrophages, however, was dramatically reduced with Ca^{2+} chelation of serum (Figure 4B).

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

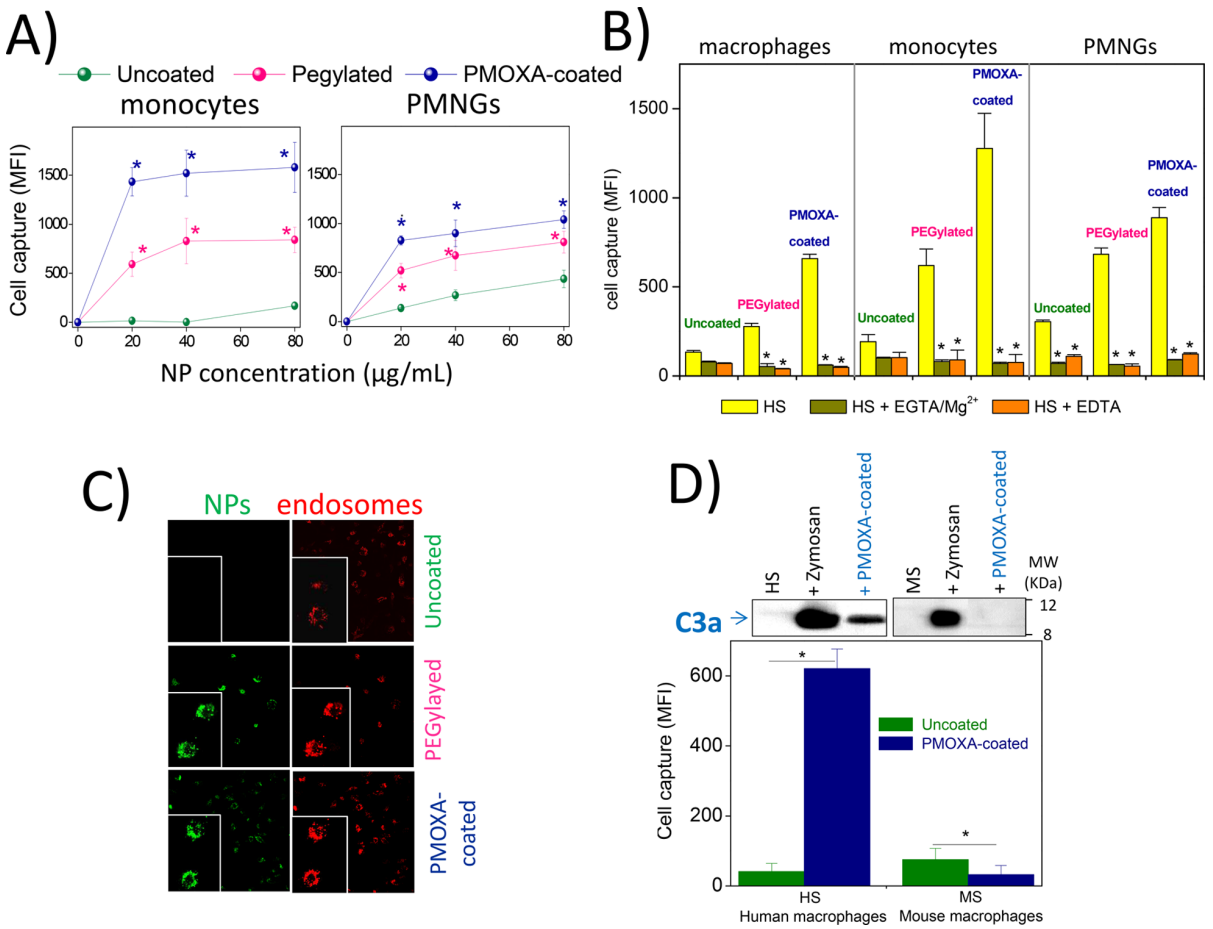


Figure 4. Capture of PMOXA-coated NPs by human phagocytes. (A) Uncoated and polymer-coated NPs were treated with HS and incubated at designated concentrations with primary human blood monocytes and PMNGs. After being incubated for 3 h at 37 °C, cells were washed with PBS and their mean fluorescence intensity (MFI) was measured by FACS as a read out of NP−cell association (instrumental setting was constant to ensure cell−cell quantitative comparison). Data are mean ± SE (n = 3; samples in triplicate in each experiments). (B) NP uptake by human macrophages and blood leukocytes following NP treatment in untreated or chelated HS. NP concentration was 80 μg/mL (n = 4; samples in duplicate in each experiments); *statistical significance ($p < 0.05$) compared with respective controls (HS). (C) Fluorescent microscopy images of NP internalization by human monocyte-derived macrophages. NP concentration was 80 μg/mL. Green and red correspond to NPs and Lysotracker-labeled acidic endolysosomes, respectively. Insets are magnified views of selected fields. (D) Species difference in PMOXA-coated NP-mediated complement activation and macrophage uptake. PMOXA-coated NPs were incubated for 30 min at 37 °C with either 75% (v/v) human serum or mouse serum. Fluid-phase C3a was measured by Western blot (top panels), indicating NP-mediated complement activation. Negative controls (sera with no NPs) and positive controls (Zymosan-treated sera) were also run and analyzed as indicated. After sera incubation, NPs were diluted 10-fold in RPMI medium and further incubated for 3 h at 37 °C with human or mouse primary macrophages (as indicated). Uncoated NPs were also used in this assay for comparison. NP capture was estimated by FACS and expressed as MFI. Data are the mean ± SE (n = 3, samples run in triplicate); *$p < 0.05$.

This observation may suggest a likely role for complement activation and surface C3 fixation in NP recognition by human phagocytic cells. In contrast to human phagocytes, PMOXA-coated NPs were not only poor activators of the mouse complement system but also poorly recognized by mouse monocyte-derived macrophages compared with uncoated NPs (Figure 4D).

Next, we investigated whether complement activation and surface C3 fixation play a critical role in NP recognition by human phagocytes. Proteomic studies did not show significant surface association of collectins and associated proteases (*e.g.*, mannose binding lectin (MBL), ficolins, collectin 11, and MBL-associated serine proteases, MASPs)[38,39] with NPs. First, this observation excludes the potential involvement of these molecules and hence the lectin pathway in complement activation and C3 fixation. However, on the basis of the observed C2 discrepancy between shot-gun proteomics and functional complement activation studies, we sought to further investigate whether the lectin pathway of the complement plays a role in NP recognition by macrophages. There was no effect of mannose and *N*-acetyl glucosamine (which compete with the binding of MBL and ficolins to their substrates)[8,39] on NP uptake (Figure S9). Furthermore, aprotinin (a MASP inhibitor)[40] did not affect the capture of PMOXA-coated NPs by macrophages (Figure S10). Following these studies, we measured macrophage uptake of NPs after incubation with HS depleted of various complement proteins (Figure 5). The results showed that the recognition of polymer-coated NPs by macrophages is blocked by >95% on C3 depletion, thereby suggesting a role for C3b/iC3b as the major opsonic molecule and of cellular complement receptors 3 and 4 (Figure 5A). In addition, depletion from HS of either C1q or C4 (which are required for assembly of the classical pathway C3 convertase) or factor B (which is required for

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

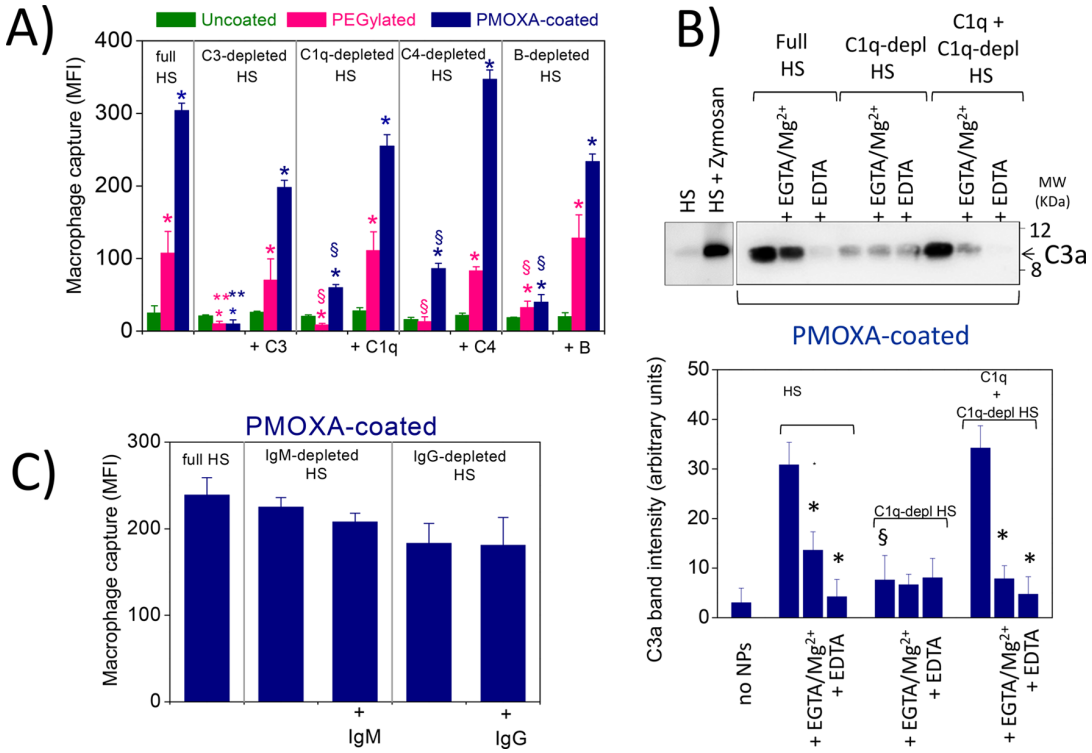


Figure 5. Mechanism of complement-mediated PMOXA-coated NP macrophage capture. (A) Macrophage uptake (3 h, 37 °C) of uncoated and polymer-coated NPs treated with control HS (full HS) or sera depleted of selected complement proteins (C1q, C3, C4, and factor B). Uptake was further monitored on restoration of the depleted factor; *statistical significance ($p < 0.05$) compared with corresponding uncoated NPs; §statistical significance ($p < 0.05$) compared with corresponding incubations in complement factor-depleted and complement factor-restored sera. (B) PMOXA-coated NPs were treated as in (A) with control HS (full HS), C1q-depleted HS, and C1q-depleted HS plus purified C1q (100 μg/mL) for 30 min at 37 °C in the absence or the presence of EGTA/Mg^{2+} or EDTA, as indicated. C3a was then measured in the incubation media by Western blot (a representative blot is shown in the top panel, whereas densitometric evaluation is depicted in the bottom panel); *statistical significance ($p < 0.05$) compared with HS in each condition; §statistical significance ($p < 0.05$) compared with C1q-depleted HS serum as well as with C1q restoration. (C) Efficacy of PMOXA-coated NP uptake by macrophages after incubation with normal HS (full HS) and with the same HS after IgM or IgG depletion. Purified autologous IgM and IgG were added to the respective depleted sera for comparison. Macrophage uptake was measured as mean fluorescent intensity (mean ± SE, $n = 3$, incubations in duplicate). Differences were not statistically significant ($p > 0.05$).

increasing the turnover of the alternative pathway, and the amplification loop) strongly inhibited (77−86%) capture of PMOXA-coated NPs by macrophages (Figure 5A). Addition of depleted factors to respective sera restored macrophage capabilities in capturing polymer-coated NPs and comparable to levels seen in intact serum (Figure 5A). These observations confirm the role of complement activation and C3 surface fixation in NP capture by macrophages. Indeed, the results in Figure 5B confirm the notion that restoration of C1q in C1q-depleted serum triggers complement activation through the classical pathway and subsequent C3 cleavage.

Antibodies of both IgG and IgM classes are known to trigger activation of the classical pathway by facilitating C1q docking. Accordingly, complement activation by PMOXA-coated NPs may have been initiated through binding of either nonspecific or putative "PMOXA-specific" antibodies in human serum. Selective depletion of either IgM or IgG from HS did not significantly affect macrophage capture of PMOXA-coated NPs (Figures 5C and S11). Therefore, C1q-mediated complement activation by PMOXA-coated NPs is apparently antibody-independent. Subsequently, we showed that purified human C1q binds directly to PMOXA-coated NPs but not to uncoated NPs (Figure S12) with nanomolar affinity (K_d = 11 ± 1 nM; maximal binding = 12 ± 1 C1q molecules/NP). This is rather comparable with the affinity of C1q for some of its natural ligands such as DNA (K_d = 22 nM), histones, annexins, and IgG (K_d range = 1−2 nM)[41] and phosphatidylserine (K_d = 37−70 nM).[42] For a comparative purpose, we also showed low affinity of pooled human IgG and lipoproteins (*e.g.*, HDL) for PMOXA-coated NPs (Table 1 and Figure S13).

Table 1. Binding Constants (K_d) and Maximal Binding Stoichiometry (B_{max}) of the Indicated Purified Human Purified Proteins to PMOXA-Coated NPs[a]

species	K_d ± SE (nM)	B_{max} (number/NP)
C1q	11 ± 1	12 ± 1
IgG	1671 ± 224	48 ± 4
HDL	612 ± 344	226 ± 58

[a] K_d was calculated by nonlinear fitting, based on free protein determination in the presence of PMOXA-coated NPs at equilibrium (C1q and HDL) or on NP-recovered protein after washings (IgG). Values are the mean ± SE ($n = 3$; samples run in triplicate).

The globular head of C1q is predominantly basic, which makes C1q function as a charge pattern recognition molecule.[43] Indeed, direct binding of the highly cationic C1qA chain to anionic

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

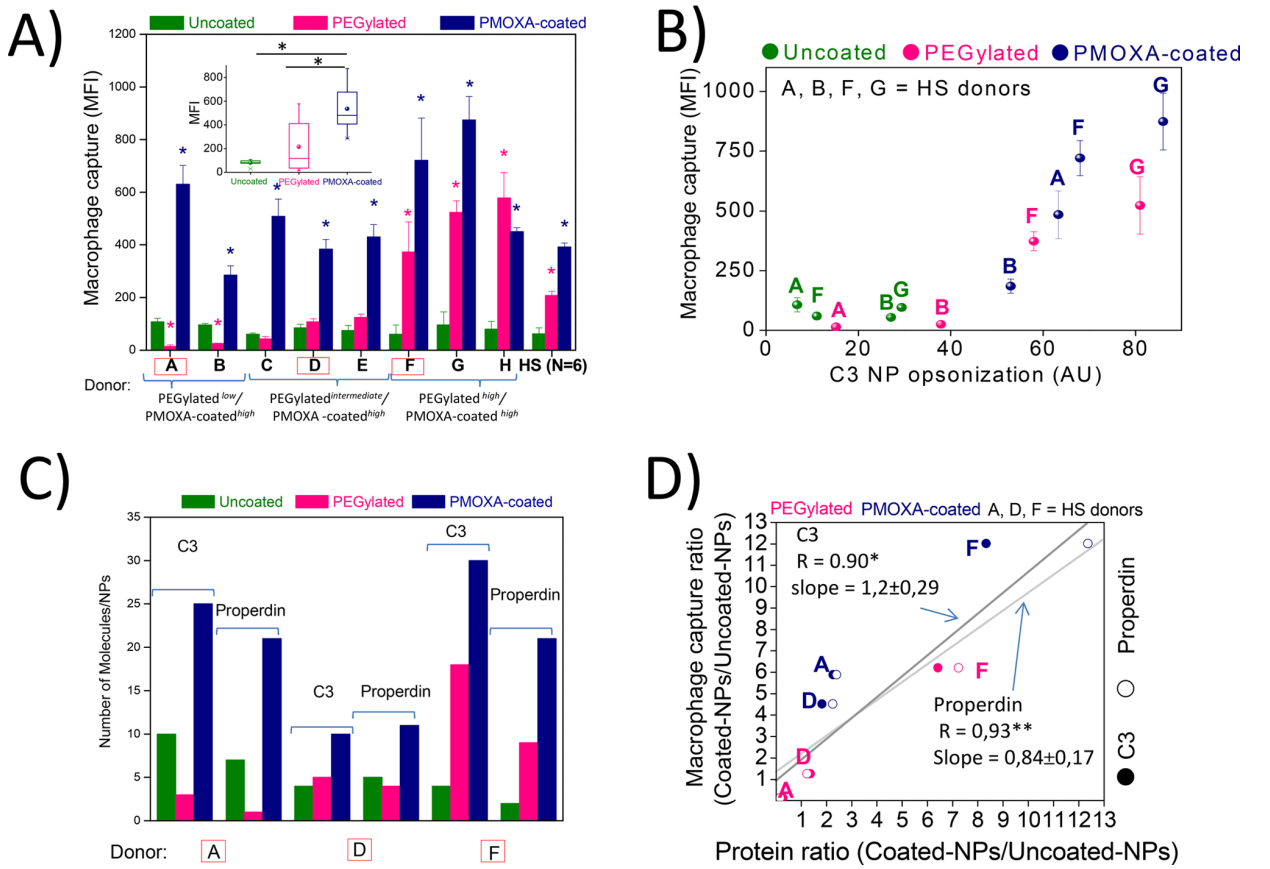


Figure 6. Interindividual variability of HS on NP capture by macrophages and complement protein association. (A) Capture intensity of uncoated and polymer-coated NPs (80 μg/mL) by human macrophages was measured by FACS after 3 h incubation at 37 °C. NPs were previously incubated with HS obtained from eight different donors (A–H) or with an independent commercially available pooled HS (CompTech, HS). The results are mean ± SE (n = 5; duplicate incubations); *statistical significance compared with uncoated NPs in the same HS sample ($p < 0.05$). The inset represents the data from different donors as a *whisker plot* (bold symbol = mean; box = ±SE; whiskers = max and min values) to show the general trend and the statistical variability (*$p < 0.05$). HS was ranked in three groups based on their efficacy to mediate PEGylated NP capture compared with uncoated NPs: PEG^{low} (PEGylated NPs capture statistically lower compared with uncoated NPs); PEG^{intermediate} (PEGylated NPs capture statistically not different from uncoated NPs); PEG^{high} (PEGylated NPs capture statistically higher than that of uncoated NPs). The red-boxed HS donors represent these three categories and were used for shot-gun proteomics. (B) Comparison of macrophage capturing efficacy of NPs after preincubation with HS from four donors (A, B, F, and G), measured by FACS after 3 h of incubation and represented as MFI (mean ± SE; n = 4, samples in duplicate). The quantity of βme-released corona C3 β subunit (arbitrary units) was determined by densitometry following SDS-PAGE. (C) Stoichiometry of C3 and properdin based on iBAQ analysis following NP incubation with HS from the indicated donors and shot-gun proteomics. (D) Macrophage capture efficacies of polymer-coated NPs preincubated in sera of three donors (A, D, and F) were normalized to that of uncoated NPs one (macrophage capture ratio) and plotted against the amounts of C3 and properdin [estimated from the *intensity* parameter after shot-gun analysis, as well normalized to uncoated NPs values (protein ratio)]. Pearson correlation index (R) and the slope ± SE of the linear regression interpolation for both proteins are indicated (*$p < 0.05$; **$p < 0.01$).

liposomes (*e.g.*, cardiolipin-containing vesicles) was demonstrated previously, which led to complement activation.[44] In addition to this, C1q was also suggested to bind cardiolipin-containing liposomes and PEGylated vesicles through hydrophobic interactions and hydrogen bonding.[44] Given the fact that the ζ-potential of PMOXA-coated NPs is close to neutrality, it is most likely that direct C1q association with these NPs could arise from hydrophobic interactions and/or hydrogen bonding by considering that in each PMOXA molecule there are about 40 C═O and 2 S═O functionalities, and that S═O groups are buried close to NP surfaces, whereas C═O groups are distributed along the chains. It is also plausible that in serum/plasma, C1q may also function as a charge pattern recognition molecule by binding to exposed anionic domains of protein corona on PMOXA-coated NPs.

Effect of Sera Variability on Complement Activation, Protein Corona, and NP Uptake by Macrophages. Next, we compared the effect of sera from 8 individual donors on NP uptake by human phagocytes to account for interindividual variations in complement opsonic activities. The results in Figure 6A show that all 8 sera promoted efficient uptake of PMOXA-coated NPs by macrophages compared with uncoated NPs. In the case of PEGylated NPs, macrophage uptake was more variable and translated either to smaller, equal, or higher values than that of uncoated NPs, depending on serum donor. C3 β chain densitometry analysis of NPs incubated with different sera (Figure 3B and Figures S7 and S14) also mirrored macrophage captures efficacies (Figure 6B). Comparative shot-gun proteomics further confirmed the above-mentioned observations, which correlated with C3 and properdin deposition on NPs, when tested in three representative sera displaying different macrophage capturing efficiency [(i) a serum that showed low macrophage capture of PEGylated NPs but high

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834–5847

levels of PMOXA-coated NPs (donor A; PEG^{low}/PMOXA^{high}), (ii) a serum promoting moderate/intermediate capture of PEGylated NPs, yet high levels of PMOXA-coated NPs (donor D; PEG ^{intermediate}/PMOXA ^{high}), and (iii) a serum promoting high capture of both coated NP types by macrophages (donor F; PEG^{high}/PMOXA ^{high})] (Figure 6C).

Furthermore, based on the label-free quantification parameter iBAQ, we observed no major quantitative differences in the protein corona for all NP types in the three sera (Figure S15, SI files 1−3 and 8−12). Subsequently, we compared the polypeptide composition of the corona (Figure S16, SI files 1−3 and 8−12). Clusterin and lipoproteins constituted major corona proteins from all sera. Although clusterin showed no clear preference for any specific NP type, Apo-AIV and Apo-E were consistently enriched on PEGylated NPs, whereas Apo-CI, CII, CIII, and Apo-AII were preferentially associated with PMOXA-coated NPs. Uncoated NPs effectively bound all indicated apolipoproteins but with no clear selectivity. Finally, we statistically correlated the macrophage capture efficacy of polymer-coated NPs and the amount of those major protein classes consistently found in their corona through normalization (Figure 6D). As expected, the analysis showed a significant positive correlation between macrophage capture efficacy and the quantity of deposited C3 (Pearson's $R = 0.9$, $p < 0.05$) and properdin (Pearson's $R = 0.93$, $p < 0.01$). On the contrary, macrophage capturing efficacies of polymer-coated NPs did not significantly correlate with surface IgG (Pearson's $R = -0.17$, $p = 0.75$) and IgM binding (Pearson's $R = 0.57$, $p = 0.24$). These observations corroborate with immunoglobulin depletion studies of HS and complement activation profiling. We also found no significant correlation between NP surface enrichment of clusterin, a protein reported to exhibit dysopsonic activity,[45] and NP recognition by macrophages (Pearson's $R = -0.62$, $p = 0.18$). Clusterin is a regulator of complement activity and is incorporated stoichiometrically into terminal complement complex,[46−48] thereby inhibiting its lytic activity and proinflammatory effects.[49] In accordance with complement activation properties of the tested NPs, and clusterin-mediated regulation of the complement by accelerating sC5b-9 formation, clusterin deposition on NP surfaces seems physiologically relevant. It is therefore possible that a fraction of surface-bound clusterin is in complex form with terminal complement C7, C8 α, and C9b proteins.[49] Furthermore, it is also known that clusterin circulates in plasma as a HDL complex with ApoA-I.[50] Because clusterin deposition on NPs still occurs when the complement is inhibited (Figure 3 and Figures S5−S7), this may indirectly support the notion of direct HDL binding to NPs and/or clusterin-ApoA-I transfer from HDL to NP surfaces. Nevertheless, our data show that human macrophages can efficiently capture NPs on efficient C3 opsonization, despite NP surface enrichment with clusterin.

CONCLUSIONS

We demonstrated that NP coating with PMOXA, which confers "stealthing" properties in the murine environment, effectively activates the human complement system through C1q binding independent of initial antibody deposition. C1q-mediated complement activation, in turn, enhanced C3 opsonization through the amplification loop of the alternative pathway. The extent of both C3 and properdin deposition on PMOXA-coated NPs correlated with human macrophage capture efficacy. We further showed that efficient C3 opsonization supersedes the proposed dysopsonic effect of clusterin and other corona proteins on NP surfaces. At large, our data show good correlation between proteomics and functional immune studies (with the exception of C2 protein) and further draw attention to the importance of testing nanomaterial immune compatibility with human materials, rather than relying on non-human sources (animal sera, cells, and *in vivo* studies in the murine model). Such disparities in immune handling of nanomaterials have also been observed among rodents. For example, in mice, intravenously injected ganglioside GM1-incorporated liposomes avoid rapid interception by macrophages of the RES,[51] whereas in rats, such vesicles are not long-circulating.[52] Rats, but not mice, produce natural anti-GM1 antibodies which, by triggering the classical complement pathway, likely mediate the clearance of GM1-containing liposomes by liver macrophages. Another example is the porcine model, which has been suggested for evaluating the safety (infusion reactions) of nanopharmaceuticals under development.[16] However, a large body of evidence suggests that the porcine reactions are related to robust NP clearance by pulmonary intravascular macrophages (PIMs) and rapid release of arachidonate metabolites from these cells.[17,53] Normal human lungs, however, do not have PIMs. Accordingly, testing of nanomedicine safety in porcine (and other animals with resident PIMs) will most likely induce cardiopulmonary distress, and therefore, global nanomedicine safety assessment in such species is perhaps inappropriate.

Finally, given the versatility of oxazoline-based polymers to further chemical modifications, future attempts may be directed to synthesize derivatives with no affinity for C1q binding and push the path toward engineering of oxazoline-coated NPs that evade recognition by elements of the human innate immune system.

METHODS

Nanoparticle Preparation and Characterization. Details of all NP preparation and characterizations steps are provided in the supplementary files (S1−S4). For NTA, NPs were diluted in purified water (182 MΩ· cm) and tracks analyzed with Nanosight 2.3 software after acquisition with a LM20 NanoSight mounted with a blue (405 nm) laser (Malvern Instruments, UK). A minimum of five different videos of individual NPs were pooled to define size distribution in three independent experiments per NP type.

Cells. Blood was drawn from healthy volunteers and provided by the Blood Transfusion Centre of the Padova Hospital (Padova, Italy) complying with local regulations. Blood leukocytes were purified from buffy coats. Buffy coats were centrifuged (800*g*, 10 min, RT) and then treated twice with 155 mM NH_4Cl, 10 mM $KHCO_3$, and 100 mM Na_2EDTA at pH 7.4 for 5 min at room temperature to lyse erythrocytes. Leukocytes were then washed twice with PBS and used immediately. Forward *versus* side scatter gating in FACS allowed separation and analysis of the three different cell populations (lymphocytes, monocytes, and PMNGs). Human macrophages were derived from blood monocytes incubated for 7 days with 100 ng/mL of macrophage colony-stimulating factor (M-CSF, Peprotech, UK) in RPMI 1640 (Invitrogen, CA, USA) supplemented with 20% (v/v) fetal calf serum (Euroclone, Italy, endotoxin <0.3 EU/mL), 100 U/mL penicillin, and 100 μg/mL streptomycin).[54] Human PMNGs were purified from healthy donors as previously described using dextran sedimentation, centrifugation through Ficoll-Paque, followed by hypotonic lysis of contaminating erythrocytes and washing with PBS.[55] To obtain mouse bone marrow macrophages, femurs and tibiae of Balb/c animals, previously subjected to cervical dislocation, were flushed with Iscove's modified Dulbecco's medium (IMDM, Thermo Fisher Scientific, MA, USA). After lysis of red blood cells with ammonium−chloride−potassium lysis buffer (5 min at room temperature), leukocytes were recovered from the obtained cell suspension, and mononuclear phagocytic precursor cells were propagated in suspension in IMDM containing 10% v/v FBS (Thermo Fisher Scientific), 100 U/mL

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

penicillin, 100 μg/mL streptomycin (Thermo Fisher Scientific), and 10 ng/mL M-CSF (Miltenyi, Germany).[56] After being cultured at 37 °C in a humidified incubator with 5% CO_2 for 3 days, one-half of the medium was replaced with a fresh one, and cells were further incubated for 4 days to obtain their differentiation into macrophages.

Human and Mouse Sera. Venous blood from healthy human donors was collected into BD Vacutainer clot activator tube (Becton Dickinson, NJ, USA). Experiments with human sera were performed under local guidelines. After clotting at room temperature, blood samples were centrifuged at 1250*g* for 5 min to obtain sera. Collected sera were aliquoted and stored at −80 °C. Before use, serum was thawed at 37 °C.[57] Independent commercially pooled HS preparations from at least six donors were provided by CompTech (TX, USA) and used when indicated. C1q-, C3-, C4-, and factor-B-depleted sera and purified C1q, C3, C4, and factor B were purchased from CompTech. For IgM depletion, HS was incubated with anti-human IgM agarose beads (Sigma, MO, USA).[58] Prechilled agarose beads (0.2 mL packed) equilibrated in cold PBS were mixed with 0.2 mL of HS and 0.5 mM EDTA, and the suspension was mixed at 4 °C for 60 min. Agarose beads were pelleted by centrifugation, and the incubation step was repeated with fresh resin. IgM-depleted serum was then recovered and immediately frozen in liquid nitrogen. On the day of the experiment with macrophages, physiological Ca^{2+} and Mg^{2+} levels were restored during incubation with cells. To recover bound IgM, beads were treated with 500 μL of 100 mM glycine, pH 2.8 (Sigma), for 5 min and centrifuged. To the recovered supernatant was added 50 μL of 1 M Tris-HCl (Sigma) to restore physiological pH; this procedure was repeated three times and then dialyzed overnight at 4 °C against PBS. Finally, IgM was concentrated in Microcon tubes (Millipore, MA, USA), and protein concentration was measured by Bradford assay (Biorad, CA, USA). For the depletion of IgG antibodies, HS was incubated with protein A sepharose (Amersham, UK). To 200 μL of HS (supplemented with 0.5 mM EDTA) was added and mixed 200 μL of packed protein A beads at 4 °C for 60 min. Sepharose beads were removed by centrifugation, and the follow-up treatments and IgG recovery procedures were the same as those for IgM. Depletion of IgM and IgG from sera was confirmed by immunoblotting. Here, 0.5 μL of designated serum was mixed with 50 μL of PBS, and then 12 μL of sample containing loading buffer was loaded onto a 6 or 8% gel. Following SDS-PAGE, proteins were blotted onto nitrocellulose paper (Biorad) for 1 h. The membrane was blocked with TBS-T (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1% w/v Tween 20) containing 3% w/v nonfat milk overnight at 4 °C. Finally, the membrane was incubated with anti-IgM-HRP antibodies (Sigma) or anti-IgG-HRP antibodies (Merck, Germany) for 1 h at room temperature, and proteins were detected by a Huvitec (Eppendorf, Germany) image analyzer. Pooled mouse serum was prepared from the blood of 12 adult healthy male and female BALB/c mice. Blood clotting and retraction was allowed to take place spontaneously at room temperature following procedures outlined by Lachmann.[59] The initial centrifugation was carried out at 3000*g* for 5 min at room temperature to obtain serum. A second high-speed centrifugation was performed to remove contaminated blood cells at 20 000*g* for 2 min. Fresh serum from the second centrifugation was not frozen but was used immediately for biological experiments.

NP Binding to Cells. Intracellular distribution of fluorescent NPs was assessed by confocal microscopy. Macrophages (2×10^6) seeded on glass slides were washed with PBS (Invitrogen) and incubated for 3 h at 37 °C with NPs, preincubated with 75% (v/v) HS for 15 min at 37 °C, and then diluted in RPMI. Following incubation, the medium was removed, and cells were washed with PBS and incubated for 30 min with 75 nM LysoTracker Green DND-26 (Invitrogen) for endolysosomal staining and then directly analyzed by confocal microscopy (Leica SP2, Germany). Images were processed using ImageJ software. For FACS analysis, NPs (200−800 μg/mL) were preincubated with 75% v/v for 20 min at 37 °C. For control experiments, NPs were incubated in PBS. NPs were then diluted with RPMI medium and incubated with cells (10^6 cells/well for leukocytes and macrophages) for 3 h at 37 °C. Cells were than washed with PBS, resuspended in FACS buffer (1% FBS in PBS), and analyzed by cytofluorimetry (FACSCantoII, Becton Dickinson). Data were analyzed by FACSDiva software (BD) and expressed as MFI (mean fluorescence intensity) values, normalized based on intrinsic quantum yield of different NPs types and batches. In some experiments, the incubation of NPs with HS was performed in the presence of 10 mM EGTA/2 mM $MgCl_2$ or 10 mM EDTA to chelate Ca^{2+} or both Ca^{2+} and Mg^{2+}, respectively, before dilution as above in culture medium for cell uptake measurements. In some experiments, samples were incubated in the presence of 25 mM D-mannose, *N*-acetylglucosamine (inhibitors of lectin complement pathway) or D-galactose (as control), and 5 μM aprotinin (Sigma) and further diluted in cell culture medium in the presence of the same inhibitors.

NP Protein Corona Determination by Shot-Gun Proteomics. Uncoated, PEGylated, or PMOXA-coated NPs (800 μg/mL) were incubated under stirring at 37 °C for 15 min in RPMI 1640 supplemented with 100 μL of 75% (v/v) HS. NPs were washed three times with 1 mL of ice-cold PBS (pH 7.4) and then recovered by centrifugation (30 min, 21 000*g* at 4 °C). NP−protein pellets were resuspended in 8 M urea/50 mM NH_4HCO_3 (Sigma). Disulfide bonds were reduced with 5 mM TCEP (Sigma) for 30 min and alkylated with 40 mM iodacetamide (Fluka, Mexico) at room temperature in the dark. The content of urea was brought to 6 M with 50 mM NH_4HCO_3, and 1−100% (w/w) of LysC (Promega) was added to each sample for 6 h at 37 °C. Samples were diluted five times with 50 mM NH_4HCO_3 to obtain a urea concentration of 1.2 M during trypsin digestion. A 1−50% (w/w) of sequencing-grade-modified trypsin (Promega) was added to the protein sample, and digestion was carried out at 37 °C overnight. Samples were then centrifuged at 18 000*g* for 30 min at 4 °C to remove NPs and desalted using C18 cartridges (Sep Pak, Waters, France) according to manufacturer's instructions. All samples were treated simultaneously and exactly in the same way to allow a quantitative comparison of the MS data. Samples were then dried under vacuum and suspended in 24 μL of 3% (v/v) acetonitrile/0.1% (v/v) formic acid. Next, 2 μL of each sample was analyzed by LC-MS/MS on a LTQ-Orbitrap XL mass spectrometer (ThermoFisher Scientific) coupled online to a nano-HPLC Ultimate 3000 (Dionex−ThermoFisher Scientific) as described earlier.[60] Peptides bound to pico-frit columns (New Objective, NJ, USA) packed with C18 material (Aeris Peptide 3.6 mmXBC18; Phenomenex, CA, USA, flow rate = 8 μL/min) were separated with an acetonitrile/0.1% (v/v) formic acid gradient (from 3 to 40% in 45 min, flow rate = 250 nL/min). Ion source capillary temperature was 200 °C, and spray voltage was set to 1.2−1.3 kV. The instrument operated in a data-dependent mode with a full scan at 60 000 resolution on the Orbitrap followed by MS/MS fragmentation in the linear trap of the four most intense ions.

Protein Identification. Raw data files were processed using the software Proteome Discoverer 1.4 (ThermoFisher Scientific) interfaced to a Mascot server (version 2.2.4; Matrix Science, MA, USA) and the software MaxQuant (version 1.5.1.2)[61] integrated with the Andromeda search engine.[62] Protein identification was performed against the human section of the Uniprot database (version 20150401; 9411 sequences). Enzyme specificity was set to trypsin with a maximum of two missed cleavages allowed. Precursor and fragment mass tolerances were set to 10 ppm and 0.6 Da, respectively. Carbamidomethylation of cysteine residues and methionine oxidation were set as a static and dynamic modification, respectively. For data obtained with Proteome Discoverer, the algorithm Percolator was used to assess peptide and protein identification confidence.[63] Data were filtered to account only for proteins identified with at least two independent unique peptides with high confidence ($q \leq 0.01$). Proteins were grouped into protein families according to the principle of maximum parsimony. Label-free quantification was obtained with the precursor ion area detector node, which extracts and integrated the area under the peak for each identified peptide and estimates the amount of protein by averaging the area under the peak of identified peptides for each protein (intensity). Therefore, quantitative intensity data are associated only with proteins identified with at least three independent peptides. For data obtained with the software MaxQuant, the intensity of each peptide and the intensity-based absolute protein quantification (iBAQ) parameter (described in ref 64) were automatically calculated by the software. Data were manually filtered to keep into account only proteins identified with at least four unique peptides per identified proteins.

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

Estimation of Polypeptide Stoichiometry and Mass in NP Corona. The approximate polypeptide composition of the NP corona was calculated by taking into account the size and density of NPs (mean diameter of 100 nm and density around 1.5 g/mL, leading to a M_w of $\sim 5 \times 10^8$ Da for all NPs) and then using proteomics label-free parameter iBAQ (4 peptide cutoff). Based on the above information, a ratio of 1 clusterin molecule (49 000 Da) per NP would lead to a weight ratio of 0.01 μg clusterin/100 μg of NPs. Hence, using values of the experimentally determined percent amount of total protein (associated with 100 μg of NPs) (% w protein), we used the following equation to obtain clusterin stoichiometry (number of polypeptide/NPs):

$$\text{clusterin/NP stoichiometry} = \left[\left(M_w \text{ clusterin} \times \text{iBAQ clusterin}/\sum M_{w,n}\text{iBAQ}_n\right) \times \%\text{w proteins}\right] \times 0.01^{-1}$$

The stoichiometry of other polypeptides was derived by their respective iBAQ values, compared to that of clusterin (polypeptide$_n$/NP = number of clusters/NPs $\times$ (iBAQ$_n$/iBAQ$_{\text{clusterin}}$). The mass amount was determined based on the molecular weight of a given polypeptide and the Avogadro number using the equation

$$\text{weight of one polypeptide}_n = [M_{w,n}\,(\text{KDa})/6.022] \times 10^{-20}\ \text{gr}$$

SDS-PAGE and Staining. Uncoated, PEGylated, and PMOXA-coated NPs (800 μg/mL) were incubated under stirring at 37 °C for 15 min in RPMI 1640 supplemented with 100 μL of 75% (v/v) HS. NPs were washed three times with 1 mL of ice-cold PBS, pH 7.4, and recovered by centrifugation (30 min, 21 000g at 4 °C). The NP−protein pellet was dissolved in 25 μL of loading sample buffer (62.5 mM Tris-HCl, pH 6.8, 2% w/v SDS, 25% v/v glycerol, 0.01% w/v bromophenol blue, with or without β-mercaptoethanol). After being heated at 95 °C for 5 min, equal volumes (12 μL) of samples were subjected to SDS-PAGE in 8 or 12% acrylamide gels. After being heated at 95 °C for 5 min, equal volumes (12 μL) of samples were subjected to SDS-PAGE (8 or 12% acrylamide). NP-independent protein recovery was assessed by running mock samples in protein media. Silver or Coomassie G-250 staining (for mass spectrometry) was used to reveal separated polypeptides. For silver staining, gels were fixed for 30 min in 50% v/v methanol 10% v/v acetic acid and then incubated for 15 min in 5% v/v methanol, 1% v/v acetic acid, washed three times with water, and exposed for 90 s to thiosulfate solution (200 μg/mL $Na_2S_2O_3$ pentahydrate). After extensive washing with water, gels were incubated in the dark for 30 min with 0.2 g/L $AgNO_3$, rinsed, and developed for 5−15 min with a solution containing 60 mg/mL Na_2CO_3, 4 μg/mL $Na_2S_2O_3$ pentahydrate, and 0.01875% v/v formaldehyde. Reaction was stopped with 6% v/v acetic acid. For colloidal Coomassie G-250 staining, gels were fixed for 18 h in 50% v/v methanol and 2% v/v H_3PO_4, rinsed three times for 30 min with water, and incubated for 1 h in a solution containing 34% v/v methanol, 2% v/v H_3PO_4, and 17% w/v $(NH_4)_2SO_4$. Staining was performed for 3 days in 34% v/v methanol, 2% v/v H_3PO_4, 17% w/v $(NH_4)_2SO_4$, and 0.066% w/v Coomassie G-250, and it was followed by destaining in water for an additional 3 days. Band densitometry was performed using ImageJ software, after background subtraction. In separate experiments, the protein content after corona formation was determined by Bradford assay after washings. Bovine serum albumin (Sigma) was used as the reference protein.

Protein Identification after In-Gel Digestion. Bands from SDS-PAGE were washed with 50% v/v acetonitrile (ACN) in 0.1 M NH_4HCO_3, vacuum-dried, treated for 30 min at 56 °C with 10 mM DTT in 0.1 M NH_4HCO_3, cooled at 25 °C, and further incubated in the dark for 20 min with 55 mM iodacetamide in 0.1 M NH_4HCO_3 to alkylate thiols. After being washed with 50% ACN in 0.1 M NH_4HCO_3, the dried gel pieces were swollen and treated overnight at 37 °C with 15 μL of digestion buffer (12.5 ng/μL trypsin, Promega, in 25 mM NH_4HCO_3). The obtained peptides, extracted as detailed by Kim *et al.*,[65] were analyzed by LC-MS/MS on a 6520 Q-TOF mass spectrometer (Agilent Technologies) coupled to a chip-based chromatographic interface. Large capacity chips (C18, 150 μm × 75 μm) and enrichment columns (C18, 9 mm, 160 nL volume, 0.3 μL/min) were employed for peptide separation. Eluting solutions A (water/formic acid 0.1% v/v) and B (acetonitrile/formic acid 0.1% v/v) were used. A gradient of solution B from 5 to 50% in 20 min allowed effective chromatographic resolution. Raw data, converted into mascot generic format with MassHunter qualitative analysis software version B.03.01 (Agilent Technologies), were analyzed with Mascot Search Engine version 2.2.4 (Matrix Science). MS/MS spectra were screened *versus* SwissProt database (June 2014 version, Taxonomy Mammalia, 66370 peptide entries). Enzyme specificity was set to trypsin/P with two missed cleavage (mass tolerance 20 ppm for parent mass and 0.6 Da for fragment ions). Cysteine carbamidomethylation and methionine oxidation were set as fixed or variable modifications, respectively. Positivity cut off: more than 1 peptide per protein ($p < 0.05$).

Western Blot, Dot Blot, and ELISA Assays of Complement Proteins and Fragments. *Western Blot.* To control complement activity, 25 μL of HS was treated with 6.25 μL of Zymosan (25 mg/mL, Sigma, prepared as described by manufacturer's instructions) at 37 °C. The reaction was terminated after 30 min with 25 mM EDTA. To assess complement activation induced by NPs, 25 μL of HS was incubated with 800 μg/mL of different NP preparations for 30 min at 37 °C. Then 1.6 μL of each sample was mixed with 38.4 μL of water and 6.7 μL of loading buffer without β-mercaptoethanol, and 15 μL of sample was loaded onto an 8 or 12% gel. Polypeptides were then transferred to nitrocellulose membranes (Biorad) which, after saturation with TBS-T containing 1% nonfat milk at room temperature for 1 h, were further treated in the same solution with antibodies against C3a or B factor (Calbiochem, 1:2000) overnight at 4 °C. Membranes were washed three times with TBS/0.1% v/v Tween 20 and treated with secondary HRP-(Calbiochem) or FITC (Millipore)-conjugated Abs. Proteins were detected by Huvitec (Eppendorf) or Odissey CLx (Li-cor) imaging systems.

Dot Blot. For the binding assay of complement proteins, NPs were treated with HS at a 1:3 (v/v) ratio (routinely mixing 10 μL of NPs 3 mg/mL and 30 μL of HS) with or without 10 mM EDTA, 10 mM EGTA/Mg^{2+}, or antiproperdin antibodies (10 folds diluted). NP concentrations were quantified with NanoSight (Malvern Instruments). After being washed three times by centrifugation in PBS plus Ca^{2+}/Mg^{2+} at 4 °C (Beckman Optima TLX ultracentrifuge, 150 000g), NP pellets were resuspended in 20 μL of PBS, and bound proteins were detached by 2% SDS and quantified. Two microliter triplicate sample aliquots were blotted on a 0.2 μm pore nitrocellulose membrane (Bio-Rad) which was then saturated with 5% (w/w) nonfat dry milk in PBS-T (1× PBS with 0.1% Tween 20) for 1 h at room temperature. Membranes were incubated with anticomplement protein specific antibodies for 1 h at room temperature, washed three times with PBS-T, and eventually challenged for 1 h with IRDye 800CW-labeled secondary antibodies. Intensities of scanned bot images, obtained with an Odyssey infrared imager (Li-COR Biosciences, Lincoln, NE, US), were measured using ImageJ software.

ELISA. Complement activation in HS was quantified by testing the NP-induced rise of complement activation products C5a, Bb, and sC5b-9 with Quidel (Quidel, CA, USA) kits as detailed.[17,66] NPs were added to 75% (v/v) HS in Eppendorf tubes (in triplicate) in a shaking water bath at 37 °C for 30 min. Reactions were terminated by addition of "sample diluent" provided with the assay kit or saline containing 25 mM EDTA. NP-induced rises of serum complement activation products were then measured following reaction termination with "sample diluent" from kits or saline supplemented with 25 mM EDTA and NP removal. Background complement activation was estimated by incubating HS with saline. In some cases, the assay was performed in the presence of 10 mM EGTA/2.5 mM $MgCl_2$ or antiproperdin antibodies or of Zymosan as a positive control.

Statistical Analysis. Mean difference significances (0.05 level) were assessed by two samples or paired t test, when appropriate, using the Microcal Origin 8 software, after checking the Gaussian distribution (Shapiro-Wilk test, 0.05 level). Pearson correlation coefficient, R, its significance (t test), and linear regression analysis parameters were calculated using the Excell2000 software statistics package. The number of independent experiments and mean values ± SE are indicated where appropriate.

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsnano.8b01806.

NP synthesis, additional physicochemical characterization; raw shot-gun data of protein and peptide identification in NP corona in HS (donor A) before and after treatment with chelating agents, iBAQ of identified polypeptides in NP corona, and SDS-PAGE gels after *in situ* digestion for different NP types; additional information on cytotoxicity, lectin pathway inhibition, antibody depletion of serum and C1q binding experiments; raw data on densitometry and further proteomics in selected sera; supplementary methods (PDF)

SI File 1: Donor A_Protein groups_MaxQuant (XLSX)

SI File 2: Donor A_Results MQ_Peptides (XLSX)

SI File 3: Shot gun elaborated data for effect of Ca^{2+} and Mg^{2+} on ORMOSIL-NP corona (XLSX)

SI File 4: PMOXA samples w or w-o EGTA-Mg or EDTA (XLSX)

SI File 5: Uncoated NPs_MS (XLSX)

SI File 6: Pegylated NPs_MS (XLSX)

SI File 7: PMOXA-coated NPs_MS (XLSX)

SI File 8: Donor D_Protein groups_MaxQuant (XLSX)

SI File 9: Donor D_Results MQ_Peptides (XLSX)

SI File 10: Donor F_Protein groups_MaxQuant (XLSX)

SI File 11: Donor F_Results MQ_Peptides (XLSX)

SI File 12: Shot gun elaborated data for NP corona composition estimation for three HS donors (XLSX)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: seyed.moghimi@ncl.ac.uk.

*E-mail: emanuele.papini@unipd.it.

ORCID

Dmitri Simberg: 0000-0002-5288-6275

Edmondo M. Benetti: 0000-0002-5657-5714

Seyed Moein Moghimi: 0000-0003-0836-926X

Fabrizio Mancin: 0000-0003-0786-0364

Emanuele Papini: 0000-0001-6033-4473

Present Addresses

▽Institut de microbiologie, CHUV, Rue du Bugnon 48, CH-1011 Lausanne, Switzerland.

▼School of Biological Science, College of Science and Engineering, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom.

Author Contributions

R.T., L.G., E.L., C.F., S.V., P.P.L., G.A., A.G.-S., F.C., G.M., E.M.B., and L.W. did experiments. All coauthors discussed and analyzed data. F.M., E.P., D.S., and S.M.M. wrote the paper with contribution from all coauthors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Centro Trasfusionale of the Hospital of Padua (ULSS 16) for providing human blood buffy coats, and Dr. Fabio Munari of the Venetian Institute for Molecular Medicine (VIMM) in Padova for helping in murine macrophage preparations. Part of the data presented was obtained by A.G.-S. in partial fulfilment of the requirements for her Ph.D. degree in Biomedical Sciences at the University of Padova. This work was supported by the University of Padova (Ex 60% and DOR, 2014-2015-2016-2017-2018), and Strategic Project NAMECA. S.M.M. acknowledges financial support by the International Science and Technology Cooperation of Guangdong Province (reference 2015A050502002) and Guangzhou City (reference 2016201604030050) with RiboBio Co, Ltd., China. L.-P.W. acknowledges financial support from Drug Discovery Pipeline of Guangzhou Institutes of Biomedicine and Health.

REFERENCES

(1) Moghimi, S. M.; Hunter, A. C.; Andresen, T. L. Factors Controlling Nanoparticle Pharmacokinetics: An Integrated Analysis and Perspective. *Annu. Rev. Pharmacol. Toxicol.* **2012**, *52*, 481−503.

(2) Moghimi, S. M.; Parhamifar, L.; Ahmadvand, D.; Wibroe, P. P.; Andresen, T. L.; Farhangrazi, Z. S.; Hunter, A. C. Particulate Systems for Targeting of Macrophages: Basic and Therapeutic Concepts. *J. Innate Immun.* **2012**, *4*, 509−528.

(3) Salvati, A.; Pitek, A. S.; Monopoli, M. P.; Prapainop, K.; Bombelli, F. B.; Hristov, D. R.; Kelly, P. M.; Aberg, C.; Mahon, E.; Dawson, K. A. Transferrin-Functionalized Nanoparticles Lose Their Targeting Capabilities When a Biomolecule Corona Adsorbs on the Surface. *Nat. Nanotechnol.* **2013**, *8*, 137−143.

(4) Maiolo, D.; Del Pino, P.; Metrangolo, P.; Parak, W. J.; Baldelli Bombelli, F. Nanomedicine Delivery: Does Protein Corona Route to the Target or off Road? *Nanomedicine (London, U. K.)* **2015**, *10*, 3231−3247.

(5) Suk, J. S.; Xu, Q.; Kim, N.; Hanes, J.; Ensign, L. M. PEGylation as a Strategy for Improving Nanoparticle-Based Drug and Gene Delivery. *Adv. Drug Delivery Rev.* **2016**, *99*, 28−51.

(6) Gabizon, A.; Shmeeda, H.; Barenholz, Y. Pharmacokinetics of Pegylated Liposomal Doxorubicin: Review of Animal and Human Studies. *Clin. Pharmacokinet.* **2003**, *42*, 419−436.

(7) Moghimi, S. M.; Hunter, A. C.; Dadswell, C. M.; Savay, S.; Alving, C. R.; Szebeni, J. Causative Factors behind Poloxamer 188 (Pluronic F68, Flocor)-Induced Complement Activation in Human Sera. A Protective Role against Poloxamer-Mediated Complement Activation by Elevated Serum Lipoprotein Levels. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2004**, *1689*, 103−113.

(8) Hamad, I.; Hunter, A. C.; Szebeni, J.; Moghimi, S. M. Poly(ethylene Glycol)s Generate Complement Activation Products in Human Serum through Increased Alternative Pathway Turnover and a MASP-2-Dependent Process. *Mol. Immunol.* **2008**, *46*, 225−232.

(9) Armstrong, J. K.; Hempel, G.; Koling, S.; Chan, L. S.; Fisher, T.; Meiselman, H. J.; Garratty, G. Antibody against Poly(ethylene Glycol) Adversely Affects PEG-Asparaginase Therapy in Acute Lymphoblastic Leukemia Patients. *Cancer* **2007**, *110*, 103−111.

(10) Zhao, Y.; Wang, L.; Yan, M.; Ma, Y.; Zang, G.; She, Z.; Deng, Y. Repeated Injection of PEGylated Solid Lipid Nanoparticles Induces Accelerated Blood Clearance in Mice and Beagles. *Int. J. Nanomed.* **2012**, *7*, 2891−2900.

(11) Suzuki, T.; Ichihara, M.; Hyodo, K.; Yamamoto, E.; Ishida, T.; Kiwada, H.; Ishihara, H.; Kikuchi, H. Accelerated Blood Clearance of PEGylated Liposomes Containing Doxorubicin upon Repeated Administration to Dogs. *Int. J. Pharm.* **2012**, *436*, 636−643.

(12) Ma, Y.; Yang, Q.; Wang, L.; Zhou, X.; Zhao, Y.; Deng, Y. Repeated Injections of PEGylated Liposomal Topotecan Induces Accelerated Blood Clearance Phenomenon in Rats. *Eur. J. Pharm. Sci.* **2012**, *45*, 539−545.

(13) Le, Y.; Toyofuku, W. M.; Scott, M. D. Immunogenicity of Murine mPEG-Red Blood Cells and the Risk of Anti-PEG Antibodies in Human Blood Donors. *Exp. Hematol.* **2017**, *47*, 36−47.

(14) Arima, Y.; Toda, M.; Iwata, H. Complement Activation on Surfaces Modified with Ethylene Glycol Units. *Biomaterials* **2008**, *29*, 551−560.

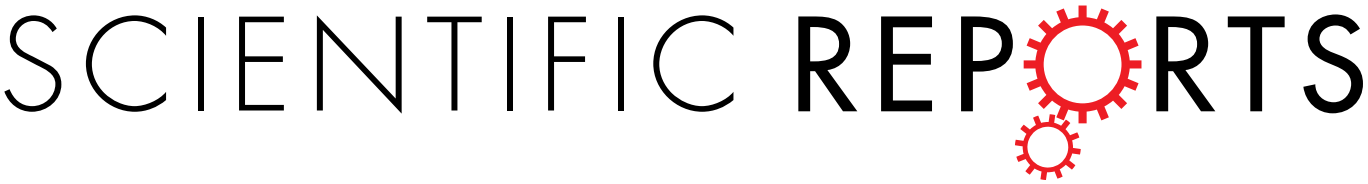
(15) Szebeni, J.; Muggia, F.; Gabizon, A.; Barenholz, Y. Activation of Complement by Therapeutic Liposomes and Other Lipid Excipient-Based Therapeutic Products: Prediction and Prevention. *Adv. Drug Delivery Rev.* **2011**, *63*, 1020−1030.

(16) Dezsi, L.; Fulop, T.; Meszaros, T.; Szenasi, G.; Urbanics, R.; Vazsonyi, C.; Orfi, E.; Rosivall, L.; Nemes, R.; Kok, R. J.; Metselaar, J. M.; Storm, G.; Szebeni, J. Features of Complement Activation-Related Pseudoallergy to Liposomes with Different Surface Charge and PEGylation: Comparison of the Porcine and Rat Responses. *J. Controlled Release* **2014**, *195*, 2−10.
(17) Wibroe, P. P.; Anselmo, A. C.; Nilsson, P. H.; Sarode, A.; Gupta, V.; Urbanics, R.; Szebeni, J.; Hunter, A. C.; Mitragotri, S.; Mollnes, T. E.; Moghimi, S. M. Bypassing Adverse Injection Reactions to Nanoparticles through Shape Modification and Attachment to Erythrocytes. *Nat. Nanotechnol.* **2017**, *12*, 589−594.
(18) Pelegri-O'Day, E. M.; Lin, E. W.; Maynard, H. D. Therapeutic Protein-Polymer Conjugates: Advancing beyond PEGylation. *J. Am. Chem. Soc.* **2014**, *136*, 14323−14332.
(19) de la Rosa, V. R. Poly(2-Oxazoline)s as Materials for Biomedical Applications. *J. Mater. Sci.: Mater. Med.* **2014**, *25*, 1211−1225.
(20) Koshkina, O.; Westmeier, D.; Lang, T.; Bantz, C.; Hahlbrock, A.; Wurth, C.; Resch-Genger, U.; Braun, U.; Thiermann, R.; Weise, C.; Eravci, M.; Mohr, B.; Schlaad, H.; Stauber, R. H.; Docter, D.; Bertin, A.; Maskos, M. Tuning the Surface of Nanoparticles: Impact of Poly(2-Ethyl-2-Oxazoline) on Protein Adsorption in Serum and Cellular Uptake. *Macromol. Biosci.* **2016**, *16*, 1287−1300.
(21) Luxenhofer, R.; Schulz, A.; Roques, C.; Li, S.; Bronich, T. K.; Batrakova, E. V.; Jordan, R.; Kabanov, A. V. Doubly Amphiphilic poly(2-Oxazoline)s as High-Capacity Delivery Systems for Hydrophobic Drugs. *Biomaterials* **2010**, *31*, 4972−4979.
(22) Konradi, R.; Acikgoz, C.; Textor, M. Polyoxazolines for Nonfouling Surface Coatings−a Direct Comparison to the Gold Standard PEG. *Macromol. Rapid Commun.* **2012**, *33*, 1663−1676.
(23) Mansfield, E. D.; Sillence, K.; Hole, P.; Williams, A. C.; Khutoryanskiy, V. V. POZylation: A New Approach to Enhance Nanoparticle Diffusion through Mucosal Barriers. *Nanoscale* **2015**, *7*, 13671−13679.
(24) Sedlacek, O.; Monnery, B. D.; Filippov, S. K.; Hoogenboom, R.; Hruby, M. Poly(2-Oxazoline)s−Are They More Advantageous for Biomedical Applications than Other Polymers? *Macromol. Rapid Commun.* **2012**, *33*, 1648−1662.
(25) Viegas, T. X.; Bentley, M. D.; Harris, J. M.; Fang, Z.; Yoon, K.; Dizman, B.; Weimer, R.; Mero, A.; Pasut, G.; Veronese, F. M. Polyoxazoline: Chemistry, Properties, and Applications in Drug Delivery. *Bioconjugate Chem.* **2011**, *22*, 976−986.
(26) Konradi, R.; Pidhatika, B.; Muhlebach, A.; Textor, M. Poly-2-Methyl-2-Oxazoline: A Peptide-like Polymer for Protein-Repellent Surfaces. *Langmuir* **2008**, *24*, 613−616.
(27) Zalipsky, S.; Hansen, C. B.; Oaks, J. M.; Allen, T. M. Evaluation of Blood Clearance Rates and Biodistribution of poly(2-Oxazoline)-Grafted Liposomes. *J. Pharm. Sci.* **1996**, *85*, 133−137.
(28) He, Z.; Miao, L.; Jordan, R.; S-Manickam, D.; Luxenhofer, R.; Kabanov, A. V. A Low Protein Binding Cationic Poly(2-Oxazoline) as Non-Viral Vector. *Macromol. Biosci.* **2015**, *15*, 1004−1020.
(29) Zhang, N.; Pompe, T.; Amin, I.; Luxenhofer, R.; Werner, C.; Jordan, R. Tailored poly(2-Oxazoline) Polymer Brushes to Control Protein Adsorption and Cell Adhesion. *Macromol. Biosci.* **2012**, *12*, 926−936.
(30) Koshkina, O.; Lang, T.; Thiermann, R.; Docter, D.; Stauber, R. H.; Secker, C.; Schlaad, H.; Weidner, S.; Mohr, B.; Maskos, M.; Bertin, A. Temperature-Triggered Protein Adsorption on Polymer-Coated Nanoparticles in Serum. *Langmuir* **2015**, *31*, 8873−8881.
(31) Woodle, M. C.; Engbers, C. M.; Zalipsky, S. New Amphipatic Polymer-Lipid Conjugates Forming Long-Circulating Reticuloendothelial System-Evading Liposomes. *Bioconjugate Chem.* **1994**, *5*, 493−496.
(32) Bludau, H.; Czapar, A. E.; Pitek, A. S.; Shukla, S.; Jordan, R.; Steinmetz, N. F. POxylation as an Alternative Stealth Coating for Biomedical Applications. *Eur. Polym. J.* **2017**, *88*, 679−688.
(33) Banda, N. K.; Mehta, G.; Chao, Y.; Wang, G.; Inturi, S.; Fossati-Jimack, L.; Botto, M.; Wu, L.; Moghimi, S. M.; Simberg, D. Mechanisms of Complement Activation by Dextran-Coated Superparamagnetic Iron Oxide (SPIO) Nanoworms in Mouse *versus* Human Serum. *Part. Fibre Toxicol.* **2014**, *11*, 62−64.
(34) Rio-Echevarria, I. M.; Selvestrel, F.; Segat, D.; Guarino, G.; Tavano, R.; Causin, V.; Reddi, E.; Papini, E.; Mancin, F. Highly PEGylated Silica Nanoparticles: "'ready to Use'" Stealth Functional Nanocarriers. *J. Mater. Chem.* **2010**, *20*, 2780−2787.
(35) Wibroe, P. P.; Ahmadvand, D.; Oghabian, M. A.; Yaghmur, A.; Moghimi, S. M. An Integrated Assessment of Morphology, Size, and Complement Activation of the PEGylated Liposomal Doxorubicin Products Doxil®, Caelyx®, DOXOrubicin, and SinaDoxosome. *J. Controlled Release* **2016**, *221*, 1−8.
(36) Sahu, A.; Lambris, J. D. Structure and Biology of Complement Protein C3, a Connecting Link between Innate and Acquired Immunity. *Immunol. Rev.* **2001**, *180*, 35−48.
(37) Merle, N. S.; Church, S. E.; Fremeaux-Bacchi, V.; Roumenina, L. T. Complement System Part I - Molecular Mechanisms of Activation and Regulation. *Front. Immunol.* **2015**, *6*, 262.
(38) Matsushita, M.; Endo, Y.; Fujita, T. Structural and Functional Overview of the Lectin Complement Pathway: Its Molecular Basis and Physiological Implication. *Arch. Immunol. Ther. Exp.* **2013**, *61*, 273−283.
(39) Wallis, R. Structural and Functional Aspects of Complement Activation by Mannose-Binding Protein. *Immunobiology* **2002**, *205*, 433−445.
(40) Petersen, S. V.; Thiel, S.; Jensen, L.; Vorup-Jensen, T.; Koch, C.; Jensenius, J. C. Control of the Classical and the MBL Pathway of Complement Activation. *Mol. Immunol.* **2000**, *37*, 803−811.
(41) Martin, M.; Leffler, J.; Blom, A. M. Annexin A2 and A5 Serve as New Ligands for C1q on Apoptotic Cells. *J. Biol. Chem.* **2012**, *287*, 33733−33744.
(42) Païdassi, H.; Tacnet-Delorme, P.; Garlatti, V.; Darnault, C.; Ghebrehiwet, B.; Gaboriaud, C.; Arlaud, G. J.; Frachet, P. C1q Binds Phosphatidylserine and Likely Acts as a Multiligand-Bridging Molecule in Apoptotic Cell Recognition. *J. Immunol.* **2008**, *180*, 2329−2338.
(43) Bradley, A. J.; Brooks, D. E.; Norris-Jones, R.; Devine, D. V. C1q Binding to Liposomes Is Surface Charge Dependent and Is Inhibited by Peptides Consisting of Residues 14−26 of the Human C1qA Chain in a Sequence Independent Manner. *Biochim. Biophys. Acta, Biomembr.* **1999**, *1418*, 19−30.
(44) Kovacsovics, T.; Tschopp, J.; Kress, A.; Isliker, H. Antibody-Independent Activation of C1, the First Component of Complement, by Cardiolipin. *J. Immunol.* **1985**, *135*, 2695−2700.
(45) Schottler, S.; Becker, G.; Winzen, S.; Steinbach, T.; Mohr, K.; Landfester, K.; Mailander, V.; Wurm, F. R. Protein Adsorption Is Required for Stealth Effect of Poly(ethylene Glycol)- and Poly(phosphoester)-Coated Nanocarriers. *Nat. Nanotechnol.* **2016**, *11*, 372−377.
(46) Choi, N. H.; Nakano, Y.; Tobe, T.; Mazda, T.; Tomita, M. Incorporation of SP-40,40 into the Soluble Membrane Attack Complex (SMAC, SC5b-9) of Complement. *Int. Immunol.* **1990**, *2*, 413−417.
(47) Bhakdi, S.; Käflein, R.; Halstensen, T. S.; Hugo, F.; Preissner, K. T.; Mollnes, T. E. Complement S-Protein (Vitronectin) Is Associated with Cytolytic Membrane-Bound C5b-9 Complexes. *Clin. Exp. Immunol.* **1988**, *74*, 459−464.
(48) Tschopp, J.; Chonn, A.; Hertig, S.; French, L. E. Clusterin, the Human Apolipoprotein and Complement Inhibitor, Binds to Complement C7, C8 Beta, and the B Domain of C9. *J. Immunol.* **1993**, *151*, 2159−2165.
(49) Hattori, R.; Hamilton, K. K.; McEver, R. P.; Sims, P. J. Complement Proteins C5b-9 Induce Secretion of High Molecular Weight Multimers of Endothelial von Willebrand Factor and Translocation of Granule Membrane Protein GMP-140 to the Cell Surface. *J. Biol. Chem.* **1989**, *264*, 9053−9060.
(50) Jenne, D. E.; Lowin, B.; Peitsch, M. C.; Böttcher, A.; Schmitz, G.; Tschopp, J. Clusterin (Complement Lysis Inhibitor) Forms a High Density Lipoprotein Complex with Apolipoprotein A-I in Human Plasma. *J. Biol. Chem.* **1991**, *266*, 11030−11036.
(51) Gabizon, A.; Papahadjopoulos, D. Liposome Formulations with Prolonged Circulation Time in Blood and Enhanced Uptake by Tumors. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 6949−6953.

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847

(52) Liu, D.; Song, Y. K.; Liu, F. Antibody Dependent, Complement Mediated Liver Uptake of Liposomes Containing GM1. *Pharm. Res.* **1995**, *12*, 1775−1780.
(53) Moghimi, S. M. Nanomedicine Safety in Preclinical and Clinical Development: Focus on Idiosyncratic Injection/infusion Reactions. *Drug Discovery Today* **2017**, DOI: 10.1016/j.drudis.2017.11.006.
(54) Fedeli, C.; Segat, D.; Tavano, R.; De Franceschi, G.; de Laureto, P. P.; Lubian, E.; Selvestrel, F.; Mancin, F.; Papini, E. Variations of the Corona HDL:albumin Ratio Determine Distinct Effects of Amorphous SiO2 Nanoparticles on Monocytes and Macrophages in Serum. *Nanomedicine (London, U. K.)* **2014**, *9*, 2481−2497.
(55) Segat, D.; Tavano, R.; Donini, M.; Selvestrel, F.; Rio-Echevarria, I.; Rojnik, M.; Kocbek, P.; Kos, J.; Iratni, S.; Sheglmann, D.; Mancin, F.; Dusi, S.; Papini, E. Proinflammatory Effects of Bare and PEGylated ORMOSIL-, PLGA- and SUV-NPs on Monocytes and PMNs and Their Modulation by F-MLP. *Nanomedicine (London, U. K.)* **2011**, *6*, 1027−1046.
(56) Stanley, E. R. Murine Bone Marrow-Derived Macrophages. *Methods Mol. Biol.* **1997**, *75*, 301−304.
(57) Fedeli, C.; Segat, D.; Tavano, R.; Bubacco, L.; De Franceschi, G.; de Laureto, P. P.; Lubian, E.; Selvestrel, F.; Mancin, F.; Papini, E. The Functional Dissection of the Plasma Corona of SiO(2)-NPs Spots Histidine Rich Glycoprotein as a Major Player Able to Hamper Nanoparticle Capture by Macrophages. *Nanoscale* **2015**, *7*, 17710−17728.
(58) Schwartz, J. T.; Barker, J. H.; Long, M. E.; Kaufman, J.; McCracken, J.; Allen, L. A. Natural IgM Mediates Complement-Dependent Uptake of Francisella Tularensis by Human Neutrophils *via* Complement Receptors 1 and 3 in Nonimmune Serum. *J. Immunol.* **2012**, *189*, 3064−3077.
(59) Lachmann, P. J. Preparing Serum for Functional Complement Assays. *J. Immunol. Methods* **2010**, *352*, 195−197.
(60) Gerotto, C.; Franchin, C.; Arrigoni, G.; Morosinotto, T. *In Vivo* Identification of Photosystem II Light Harvesting Complexes Interacting with PHOTOSYSTEM II SUBUNIT S. *Plant Physiol.* **2015**, *168*, 1747−1761.
(61) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(62) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10*, 1794−1805.
(63) Wright, J. C.; Collins, M. O.; Yu, L.; Käll, L.; Brosch, M.; Choudhary, J. S. Enhanced Peptide Identification by Electron Transfer Dissociation Using an Improved Mascot Percolator. *Mol. Cell. Proteomics* **2012**, *11*, 478−491.
(64) Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global Quantification of Mammalian Gene Expression Control. *Nature* **2011**, *473*, 337−342.
(65) Kim, S. T.; Kim, S. G.; Hwang, D. H.; Kang, S. Y.; Kim, H. J.; Lee, B. H.; Lee, J. J.; Kang, K. Y. Proteomic Analysis of Pathogen-Responsive Proteins from Rice Leaves Induced by Rice Blast Fungus, Magnaporthe Grisea. *Proteomics* **2004**, *4*, 3569−3578.
(66) Andersen, A. J.; Windschiegl, B.; Ilbasmis-Tamer, S.; Degim, I. T.; Hunter, A. C.; Andresen, T. L.; Moghimi, S. M. Complement Activation by PEG-Functionalized Multi-Walled Carbon Nanotubes Is Independent of PEG Molecular Mass and Surface Density. *Nanomedicine* **2013**, *9*, 469−473.

DOI: 10.1021/acsnano.8b01806
ACS Nano 2018, 12, 5834−5847



OPEN

Structure of an octameric form of the minichromosome maintenance protein from the archaeon *Pyrococcus abyssi*

Received: 25 May 2016
Accepted: 05 January 2017
Published: 08 February 2017

Giuseppe Cannone[1,2,3], Silvia Visentin[1,2,4], Adeline Palud[5,6,7], Ghislaine Henneke[5,6,7] & Laura Spagnolo[1]

Cell division is a complex process that requires precise duplication of genetic material. Duplication is concerted by replisomes. The Minichromosome Maintenance (MCM) replicative helicase is a crucial component of replisomes. Eukaryotic and archaeal MCM proteins are highly conserved. In fact, archaeal MCMs are powerful tools for elucidating essential features of MCM function. However, while eukaryotic MCM2-7 is a heterocomplex made of different polypeptide chains, the MCM complexes of many *Archaea* form homohexamers from a single gene product. Moreover, some archaeal MCMs are polymorphic, and both hexameric and heptameric architectures have been reported for the same polypeptide. Here, we present the structure of the archaeal MCM helicase from *Pyrococcus abyssi* in its single octameric ring assembly. To our knowledge, this is the first report of a full-length octameric MCM helicase.

Mini-chromosome maintenance (MCM) proteins are oligomeric enzymes that unwind the DNA double helix in an ATP-dependent manner. Homologues of eukaryotic MCM protein complex have been identified in all sequenced archaeal genomes, most of which have a single gene encoding for one MCM-like protein[1]. Among the exceptions, there are some Methanococcales, which possess between two and eight genes coding for MCM proteins[2,3]. In eukaryotes and archaea, MCM monomers are organized in three structural domains. The MCM amino-terminal domain (NTD) plays a role in higher order structure assembly and in the regulation of the ATPase activity, helicase activity and substrate specificity[4]. The AAA + module is needed to catalyse DNA unwinding[5]. Adjacent to the AAA + domain, the winged helix (WH) motif is proposed to have a regulative role. Many eukaryotic MCM2-7 subunits possess amino- or carboxy-terminal extensions involved in the regulation or recruitment of MCM2-7[4]. Important information on MCM helicases has been gathered from structural and functional studies of archaeal assemblies, as well as from the crystal structures of distantly related helicases (SV40, LTAg, and E1 helicase of bovine papilloma virus)[5]. A recent study of the human MCM helicase showed how this heterohexameric complex exhibits both ATP hydrolysis and DNA unwinding activities, and highlighted conformational changes on DNA binding[6]. A significant step forward in understanding the molecular organization and mechanism of the eukaryotic replicative helicase comes from the near-atomic resolution of the yeast MCM2-7 by cryo-electron microscopy. This model highlights the twisted and tilted single hexamers features of the MCM2-7 double rings and suggests a concerted mechanism for the melting of origin DNA that requires structural deformation of the intervening DNA[7].

[1]Institute of Molecular, Cell and Systems Biology, University of Glasgow, University Avenue, Glasgow G12 8QQ, UK. [2]School of Biological Sciences and Max Born Crescent, Edinburgh EH9 3JR, UK. [3]Centre for Science at extreme conditions, University of Edinburgh, Max Born Crescent, Edinburgh EH9 3JR, UK. [4]ISIS neutron source, Science and Technologies Research Council, Rutherford Appleton Laboratories, Harwell, OX11 0QX United Kingdom. [5]IFREMER, Laboratoire de Microbiologie des Environnements Extrêmes, UMR 6197, ZI de la pointe du diable CS 10070 29280 Plouzané, France. [6]Université de Bretagne Occidentale, Laboratoire de Microbiologie des Environnements Extrêmes, UMR6197, rue Dumont d'Urville 29280 Plouzané, France. [7]CNRS, Laboratoire de Microbiologie des Environnements Extrêmes, UMR6197, rue Dumont d'Urville 29280 Plouzané, France. Correspondence and requests for materials should be addressed to L.S. (email: laura.spagnolo@glasgow.ac.uk)

Archaea can provide a simplified model for understanding complex molecular machinery involved in DNA metabolism[8]. However, the differences between archaeal and eukaryotic replication machines should also be studied to provide additional information to our knowledge about this fundamental biological process, as well as to offer insights for possible biotechnological applications of unusual archaeal enzymes. The first structural observation of a multimeric assembly of MCM came from electron microscopy studies of *M. thermautotrophicus* MCM[9]. These studies showed a double hexameric structure with the two hexamers joined in a head-to-head manner, similar to the yeast MCM2-7 assembly observed later[10]. The human MCM helicase is a heterohexamer that exhibits intrinsic DNA unwinding activity[6]. MCM proteins were shown to assemble in several oligomeric states, including single hexamer, single heptamer, double heptamer and filaments[11–18]. Recently, a double octameric assembly has been reported for the ATPase domain of *Pyrococcus furiosus*[19].

Here, we describe the structural analysis of a MCM-like protein, encoded by *Pyrococcus abyssi* (Pab) genomic ORF PAB2373. We cloned, overexpressed and purified *Pab*MCM, and we analysed it by single particle electron microscopy. We also tested its helicase activity *in vitro*. Our data show that *Pab*MCM is mainly arranged as a single octameric ring, with some degree of compositional heterogeneity, and its helicase activity is weaker compared to that of *Sso*MCM. This octameric assembly is different from the other full-length MCM helicases known to date, which are mainly hexameric and sometimes heptameric.

Results

Sequence analysis of the PabMCM's AAA+catalytic domain. The sequence alignment of the archaeal MCM proteins from *S. solfataricus* (Q9UXG1), *M. thermoautrophicus* (O27798) and *P. furiosus* (Q8U3I4) with the predicted one from *P. abyssi* reveals high sequence homology (Figure S1) as well as showing the presence of two insertions that are predicted to be inteins. The first intein is inserted into the C-terminal of the Walker A motif, the second one is inserted at the N-terminus of the H2I-hp motif. *Sso*MCM and *Mth*MCM have no intein domains, while *Pfu*MCM has just one inserted before the H2I-hp motif. The N-terminal domain of *Pab*MCM reveals low sequence similarity with previously reported *S. solfataricus* and *M. Thermoautrophicus* MCM proteins (Figure S1). Biochemical and biophysical characterization demonstrated that the N-terminal domain plays a regulatory role in MCM function[20–22] as well as being involved in the formation of single and the double ring assemblies[9,23]. The low sequence similarity of the N-terminus of the A and B domains could represent a form of adaptation to the extreme environment while the higher conservation of the C domain and the NCL linker a similar role in the ring formation and inter-subunits interaction, as previously seen for other MCM protein complexes. The AAA+ catalytic domain of *Pab*MCM is better conserved compared to the N-terminus (Figure S1). In *Pab*MCM, the active site (AAA+ domain) is ~250 residues long, and it is organized in Walker A, Walker B, sensor 1 (S1) and sensor 2 (S2) motifs. Sequence alignment shows that these motifs are well conserved in *Pab*MCM although polymorphisms are present in functional motifs such as EXT-hp, Walker A, H2I-hp, PS1-hp. These differences could represent mutations evolved as a form of adaptation to the extreme habitats in which *P. abyssi* thrives.

Cloning of the intein-free PabMCM gene. The ORF PAB2373 (3336 bp) deposited at the NCBI data bank (http://www.ncbi.nlm.ni.gov) encodes a protein of 1112 amino acid residues, predicted to be a *Pab*MCM protein. The ORF contains two intein domains (Figure S1), whose insertion in the sequence cause the isolation of a fragment of the 27 amino acid residues. As result of this insertion, serine 499 is isolated from the adjacent lysine 525 of the predicted Walker A domain, while the predicted H2I-hp is not affected (Figure S1). According to the sequence alignment (Figure S1), the putative active full-length *Pab*MCM is devoid of inteins. In order to reconstruct the full-length gene, we used a PCR-based approach, in which fragments of the putative coding sequence were amplified separately and fused by PCR (Fig. 1A).

Production, DNA binding and DNA Unwinding Activity of a Soluble PabMCM Complex. *Pab*MCM was purified by heat denaturation, followed by nickel affinity and size exclusion chromatographies (Fig. 1B). To ensure that no DNA contaminant was present in the sample, fractions eluted from the nickel affinity purification step were pooled and their $A_{260/280}$ ratio measured spectrophotometerically. Only fractions with an $A_{260/280}$ between 0.51 and 0.68 (99% protein) were concentrated and loaded onto a Superose 6 10/300 GL™ size exclusion column. The apparent molecular weight of PabMCM, estimated as in ref. 24, was 575 kDa, indicating an octameric assembly.

Structural studies on archaeal MCM proteins have shown a positively charged central channel large enough to accommodate dsDNA[25,26]. In addition, biochemical studies reported that MCM proteins bind DNA through two types of structural motifs, which include a zinc finger and a beta-hairpin motif at the N-terminus and a beta-hairpin located in the AAA+ catalytic domain[22,27,28]. Bioinformatic analysis of the PabMCM sequence shows strong sequence homology and well-conserved structural motifs, with well-known archaeal MCM proteins (Figure S1).

Full-length *Pab*MCM binds dsDNA (Fig. 1C). The helicase activity of *Pab*MCM octamer complex has been examined by using a Y-shaped substrate. In the conditions tested *Pab*MCM had a mild (10% fork unwound) helicase activity, while our positive control *Sso*MCM exhibited maximal helicase activity (30% fork unwound) as previously shown[29] (Fig. 1D). Similarly, *P. furious* MCM *(Pfu*MCM) shows weak helicase activity. However, it is stimulated upon interaction by GINS complex[29]. On the other hand, the unwinding activity of the thermophilic *M. thermoautotrophicum* MCM (*Mth*MCM) is stronger than the hyperthermophilic *Pfu*MCM and *Pab*MCM enzymes[8]. Moreover, it can unwind DNA substrates despite the presence of DNA-bound proteins[30]. Like *Mth*MCM, the thermophilic *Thermoplasma acidophilum* MCM (*Tac*MCM) displays significant helicase activity[31]. Studies carried out on the mesophilic *D. melanogaster* indicated that MCM2-7 has no activity in the absence of GINS and Cdc45[28]. Based on these comparative results, we conclude that *Pab*MCM is an active helicase.

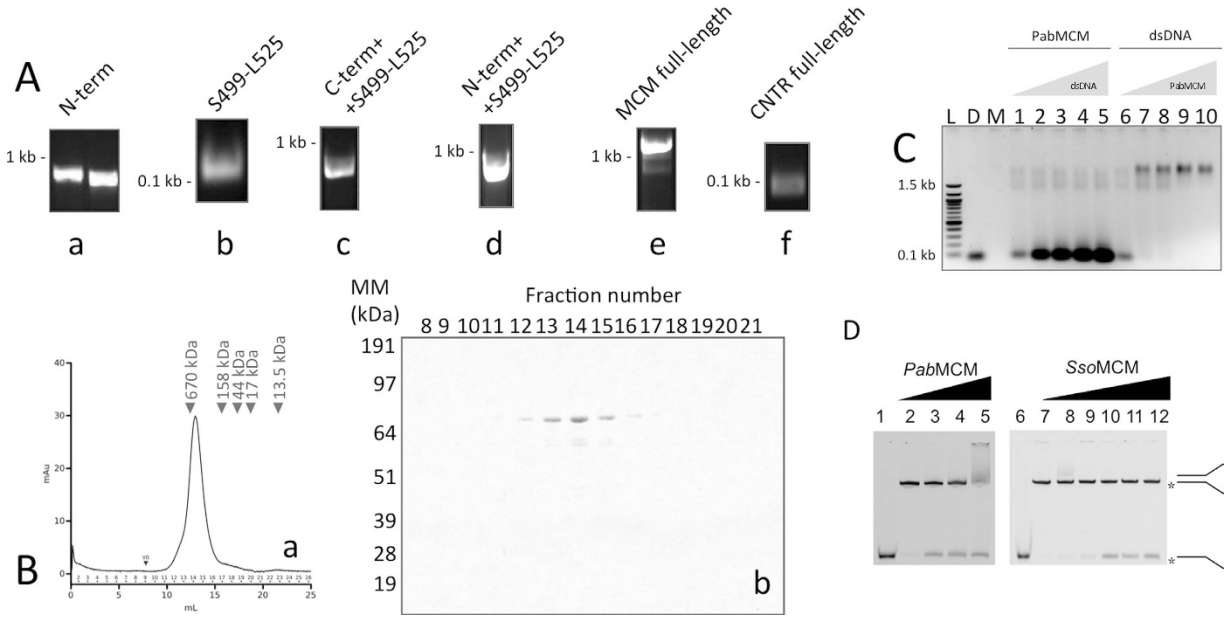


Figure 1. Cloning, purification and functional assay of PabMCM. (**A**) Reconstruction of the full-length PabMCM devoid of inteins. (a) PCR-amplified N and C-terminus. (b) PCR-amplified fragment Ser499-Leu525. (c) PCR-amplified of the joined N-term/Ser499-Leu525. (d) PCR-amplified of the joined N-term/Ser499-Leu525. (e) PCR-amplification of the *Pab*MCM full-length. (f) PCR performed in order to check that the fragment was successfully joined into the MCM sequence.(**B**) Size-exclusion chromatography. Gel-filtration was performed in 20 mM HEPES pH 7.4, 150 mM NaCl at 0.3 ml/min flow-rate. (a) Size-exclusion chromatography trace of PabMCM. The black arrow indicates the molecular weight of gel-filtration standards (BioRad). (b) SDS-PAGE analysis of eluted fractions. (**C**) DNA binding assay. The gel shows the PabMCM complex can bind dsDNA. A 59-mer dsDNA was used in this experiment. Size was chosen in order to avoid multi-MCMs loading on the same dsDNA and thus force a 1:1 stoichiometry. L, DNA ladder. D, dsDNA only used has control. M, PabMCM only used as control to avoid artefacts due to ethidium bromide staining. Lines 1–5 all contain 24 pmol (calculated on the monomeric MCM) of PabMCM with increasing amount of dsDNA (10, 20, 30, 40, 50 pmol). Lines 6–10 all contain 10 pmol of dsDNA with increasing amount of PabMCM (24, 48, 72, 96, 120 pmol). The band-shift experiment was performed as shown in Fletcher *et al.*[26]. (**D**) Helicase activity of PabMCM. Reactions contained helicase buffer with either no MCM (lanes 1 and 6, positive control or lanes 2 and 7, negative control) or increasing amounts of PabMCM (150, 300, 600 nM) lanes (3–5) and SsoMCM (50, 100, 150, 200, 300 nM) lanes (8–12).

Electron microscopy Structure of Recombinant PabMCM. We used EM coupled to single particle analysis to glean insights into the three-dimensional architecture of the *Pab*MCM complex. We calculated a three-dimensional structure of the octameric assembly using single particles acquired in negative staining conditions. A typical micrograph is shown in Fig. 2A. ~45,000 molecular images (320 × 320 pixels, 1.51 Å/pixel) were pre-processed to a final size of 80 × 80 pixels at 6.04 Å/pixel. A first round of MSA classification was performed to calculate reference-free class averages and build a first catalogue of views present in the dataset. The resulting eigenimages, calculated by MSA, are shown in Figure S2. The eigenimages are normally presented in the order of their significance[32]. In this respect, eigenimage 2 reflects the characteristic of the main shape, which is a ring, in the dataset. Additionally, the strong black ring encircling the white ring is likely to be related to the size variation between particles into the dataset[32,33]. Eigenimages 3 and 4 have two-fold symmetry, which is likely to be related to a different oligomeric states (single/double ring)[33]. Eigenimages 5 and 6 reflect size variation for the tilted views. Eigenimage 7 shows a black and a white rings, probably still related to the different particles size within the dataset[32]. To check if this interpretation was correct, molecular images were classified using the first seven eigenimages. 100 class-averages were calculated (Figure S2). Analysis of these class-averages showed three types of class-averages: one of elongated shape and two ring-shaped (Figure S2B). The validity of the interpretation of the 2nd and 7th eigenimages was tested as in White, H.E. *et al.*[32]. Rotational averages of the circular and elongated class averages were calculated and then images were compared by subtraction. As shown in Fig. 2B, both eigenimages are representative of the size variation in the dataset. Based on this observation, the dataset was partitioned into two sub-datasets: the one containing elongated particles and the one with ring-shaped ones. Further rounds of MSA classification and alignment, with a subset of ~5,000 molecular images (10% of the total) classified as elongated, revealed their features. The elongated class averages show a two-fold symmetry and a four-tier organisation, which is the typical feature of a double ring assembly of *Mth*MCM, with the top and bottom tiers corresponding to the C-termini, and the two middle tiers corresponding to the N-termini[13,15].

Analogous analyses were carried out for the 45,000-image subset of ring-shaped molecular images (90% of the total images). Further rounds of MSA classification led to classify two types of top-end views of *Pab*MCM: a small ring showing 7-fold symmetry and a bigger ring with 8-fold symmetry. The difference in size between

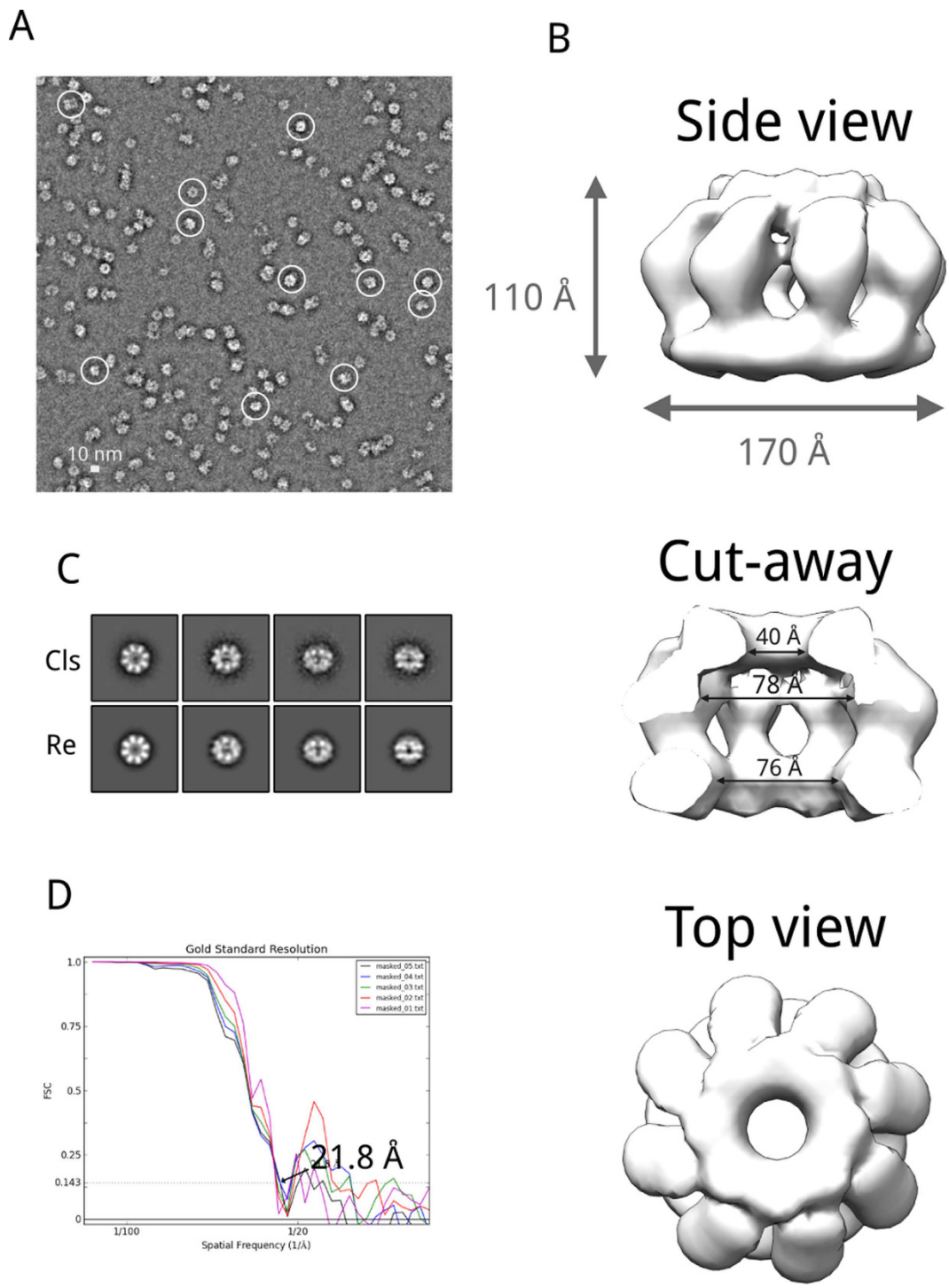


Figure 2. Electron microscopy and 3D refinement of the full–length *Pab*MCM. (**A**) Characteristic negatively stained electron micrograph of *Pab*MCM. Micrography recorded at 50,000x nominal magnification in low–dose mode (20–25 $e^{-}/Å^{2}$). White circles are used to show single particles. (**B**) Refined 3D model of the full-length *Pab*MCM single octamer. Volumes were rendered with Chimera[48]. (**C**) Class averages and re-projections for the refined 3D reconstruction. (**D**) Fourier shell correlation of the refined PabMCM model.

the two rings was 8%. MSA classification also allowed sorting two types of side views, small and large side views. More difficult was the sorting of the tilted views since the difference in size was not always clear. From this dataset, ~10,000 molecular images belonging to the 'large' *Pab*MCM assembly were taken forward for calculating the initial 3D model of the full-length *Pab*MCM using EMAN2[34,35] (Figures S3). The initial model was calculated with e2initial_model.py, which calculates a random blob model, from pure noise images, to seed a single particle reconstruction and refinement[35]. For this purpose, nine class averages were used for the 3D reconstruction of the initial model (Figure S3) and ten initial models were calculated. For each 3D reconstruction, the

e2initial_model.py program calculates a set of re-projections, which can be used to estimate the quality of the 3D reconstruction. Each 3D reconstruction was checked by inspecting the 3D model and the correspondence between the class averages and the re-projections. The model with the best matching between class averages and re-projections was refined by projection matching with e2refine.py[35]. Re-projections and class-averages were checked at each iteration. Particular care was taken for the tilted views. The final refined model is shown in Fig. 2. Projections of the maps matched with 2D class-averages assigned the same Euler angles highlighting the validity of the map (Fig. 2C). The overall shape of the refined model *Pab*MCM exhibits similarity with 3D-EM models previously reported archaeal MCMs. However, the *Pab*MCM complex exhibits an octameric assembly and has overall dimensions of 170 Å × 170 Å × 110 Å. The resolution of the map is 22 Å, calculated at 0.143 Fourier Shell Correlation (FSC) (Fig. 2D).

To test that the conditions used for negative staining did not affect the assembly of *Pab*MCM, we carried out preliminary cryo-EM experiments (Figure S4, panels A and B). We confirmed that single *Pab*MCM octamers also exist in cryo conditions. ~33,000 particles (Figure S4C) were manually picked and used to feed a 2D classification experiment in a semi-automated manner, using the Relion software[36] without imposing symmetry at any stage. The classes calculated for the whole dataset clearly show the coexistence of single ring as well as higher-order assemblies (Figure S2E). Among the single ring assemblies, we identified a majority of rings bearing 8-fold symmetry. Particles belonging to these classes were extracted from the full dataset and used as an input for reference-free 2D classification (Figure S4). These classes have features that are in excellent agreement with the ones obtained in the negative staining structure, showing that the octameric assembly is not an artifact of negative staining experiments (Fig. 2C).

Interpretation of the PabMCM 3D-EM structure. To interpret the 3D-EM reconstruction of the full-length *Pab*MCM, we docked the crystal structure of the C-terminally truncated *Sso*MCM[25]; the crystal structure of the full-length *Mka*MCM[16] and the NMR structure of the C-terminal domain of *Sso*MCM[37]. The fitting of the 3D structure of the full-length *Pab*MCM complex (Fig. 3) was performed manually using Chimera[38] and then optimised with Situs[39]. 3D volume and fitted model were rendered in Pymol[40]. The crystallographic structure of the C-terminally truncated SsoMCM fitted well into the electron density corresponding to the bottom and top tier, although a C-terminal extra density was observed (Fig. 3A and B). The orientation resulting from docking of *Sso*MCM into the map shows PS1 and HP2 β-hairpins pointing into the central channel, while Ext β–hairpin is pointing toward the side channel between the subunits composing the ring (Fig. 3C and D). Ext β-hairpin locates on the exterior side of the side channel (close-up Fig. 3D). The crystallographic structure of the full-length *Mka*MCM monomer fitted better into the electron density corresponding to the bottom and top tier (Fig. 3E and F). Docking resulted in PS1 and HP2 β–hairpins of *Mka*MCM pointing into the central channel of the 3D-EM map of the full-length *Pab*MCM (Fig. 3G and H).

The structural superposition of one octameric *Pfu*MCM AAA ring (PDB 4R7Z)[19] on the middle layer of the octameric *Pab*MCM shows that the diameter of the central cavities is very similar in both structures (Fig. 4A). Despite their different oligomeric state, the diameter of the N-terminal tier of the hexameric *PfuSso*MCM chimera (PDB 4R7Y)[19] also almost fits the top density, however it is slightly wider. On the other hand, the comparison with another octameric AAA assembly involved in DNA metabolism, the human Dmc1 recombinase (PDB 1V5W)[41] highlights that in this case the upper tier of the *Pab*MCM channel is slightly wider than that of the recombinase (Fig. 4A). The best fit for symmetry and cavity dimension could therefore be represented by a possible octameric ring of eight full-length *Mka*MCM subunits, that crystallised as monomers in 3F8T[16]. The milder activity of PabMCM compared to that of SsoMCM could be related to structural differences such as the broader end of the cavity, possibly due do the octameric arrangement of the subunits in the ring. It should however be noted that the length of each individual subunit in the octamer is not more extended than that of each of the eight modeled full-length subunits of MkaMCM (PDB 3F8T)[16]. The comparison of the hexameric *Sso*MCM and the octameric *Pab*MCM channels (Fig. 4B) shows two very different overall shapes. This different arrangement could be one explanation for the different activities of the two helicases.

Conclusions

The structural observation of archaeal MCM proteins by electron microscopy showed that several oligomeric states could be adopted by MCM proteins, including single hexamer, single heptamer, double heptamer and filaments[9–19].

In this study, we report the structure of a homo-octameric MCM protein. Electron microscopy coupled to SPA revealed that *Pab*MCM is a mixture of at least three molecular species: single heptamer, single octamer and double rings. A subset of ~10,000 particles, showing characteristic 8-fold symmetry, were classified and sorted by MSA. This subset was homogeneous enough for a subsequent 3D reconstruction. The final 3D-EM map at 22 Å (calculated by FSC) showed similarity with previously reported 3D-EM reconstruction of similar archaeal MCM proteins, however this is the first non-chimeric MCM helicase reported to assemble as an octamer.

*Pab*MCM binds to dsDNA (Fig. 1C), and has weak helicase activity compared to *Sso*MCM (Fig. 1D). Studies of another hyperthermophilic archaeal MCM, *Pfu*MCM, demonstrated that the helicase activity of MCM is stimulated by the GINS complex[42]. Like other hyperthermophilic MCM such as *Thermococcus Kodakaraensis* MCM1[2] or *Pfu*MCM, the helicase activity seems to be reduced. Additional replication factors like the GINS complex and cdc6[42,43] or post-translational modifications[44] are described to further enhance the helicase activity. Hence, similar mechanisms of activation could be involved in *Pab*MCM functionality. The structure for the full-length PfuMCM assembly is not yet available, however it is noteworthy that the octameric arrangement of PabMCM is very reminiscent of that of the PfuMCM AAA ring, suggesting an interesting similarity between two Pyrococcal proteins. The weak helicase activity could be due to the conformation of the bell-shaped inner cavity (Fig. 4B) which is broader than that of hexameric helicases such as SsoMCM (Fig. 4B).

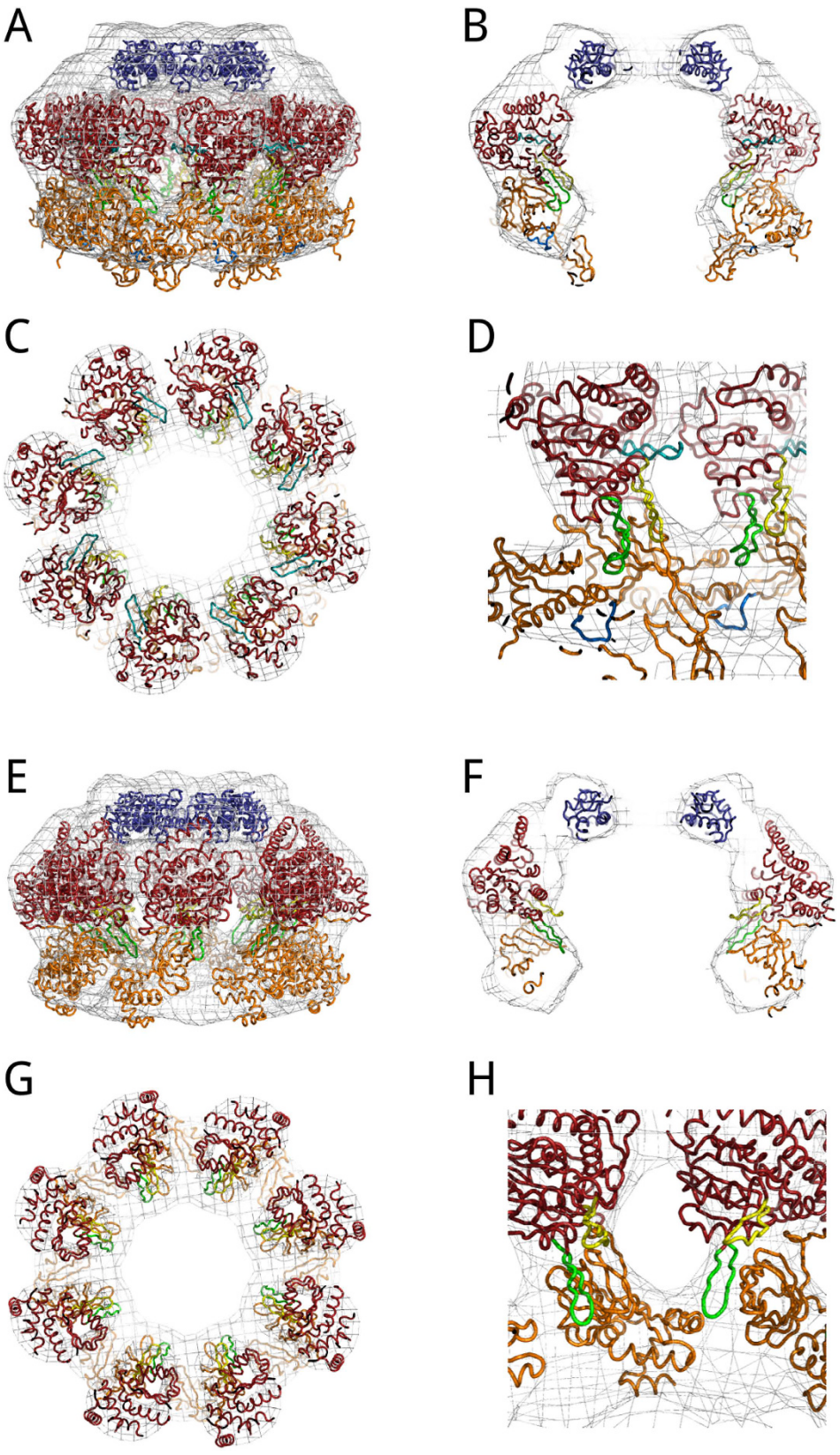


Figure 3. Model fitting of the full–length PabMCM 3D EM structure. (**A,B,C,D**) Fitting of the atomic coordinates of the C–terminally truncated *Sso*MCM (PDB 3F9V)[25] and C–terminal domain of *Sso*MCM (PDB 2M45)[37]. (**E,F,G,H**) Fitting of the atomic coordinates of full-length *Mka*MCM (PDB 3F8T)[16]. In blue is shown the fitting of the NMR structure of *Sso*MCM. In red, the AAA + module while in orange the N-terminal domain of both *Sso*MCM and *Mka*MCM. In light blue, EXT β-hairpin; in yellow, H2I β-hairpin; in green, PS1 β–hairpin; in dark blue, NT β–hairpin (only in panel D).

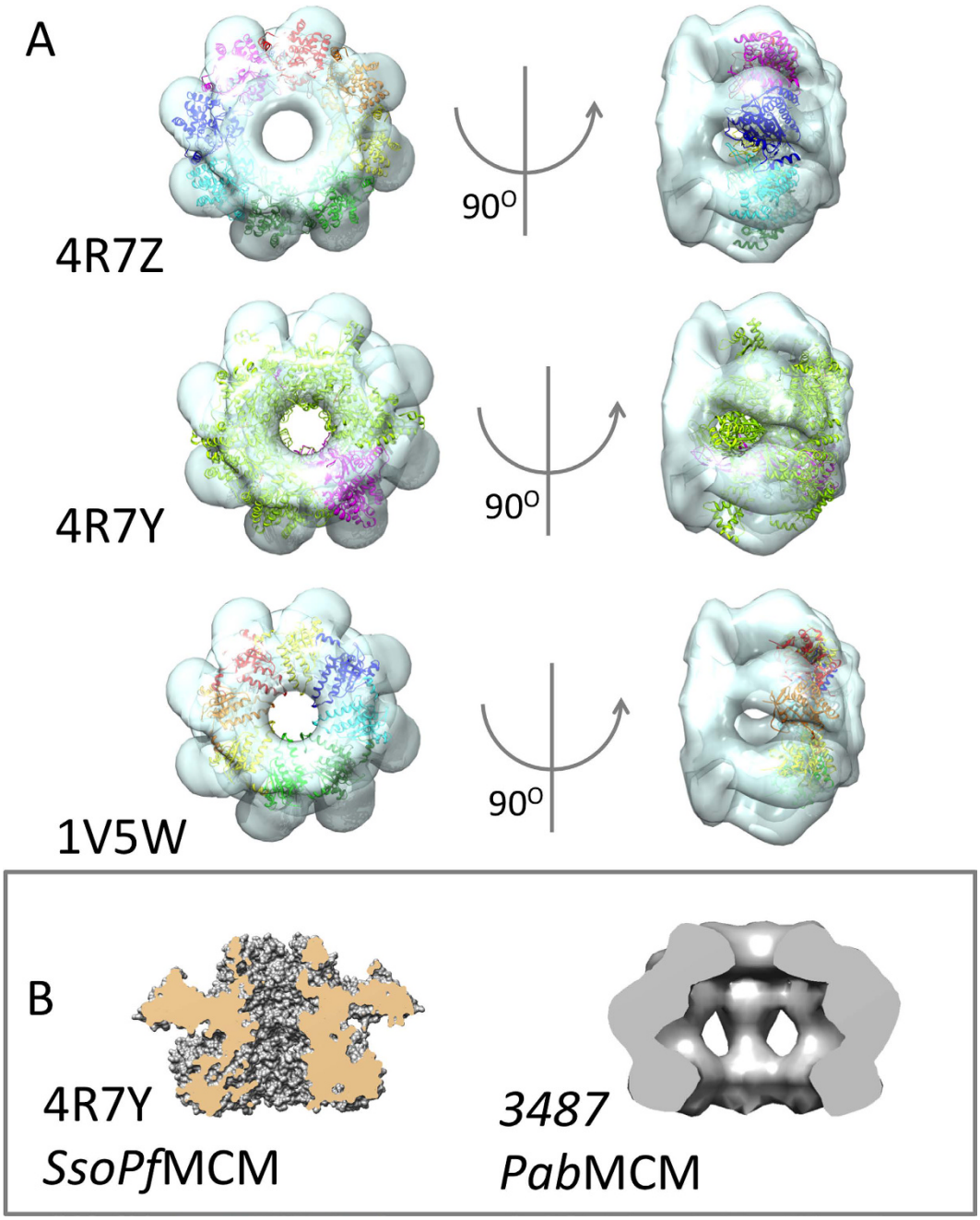


Figure 4. Structural superposition. (**A**) One octameric PfuMCM AAA ring (PDB 4R7Z)[19], the N-terminal tier of the hexameric *Pfu*Sso*MCM chimera (PDB 4R7Y)[19] and the human Dmc1 recombinase (PDB 1V5W)[41] superposed on PabMCM highlights the similarity with the symmetry and AAA ring subunits orientation of the crystallographic structure for PfuMCM. (**B**) Comparison of the central channel of the hexameric *Sso*MCM (47RY) and the octameric *Pab*MCM (EMD-3487).

Experimental Section

Cloning. The cloning was performed from Pab genomic DNA (kind gift of P. Forterre). The primers oligo1 and oligo4 were used to PCR-amplify the N-terminus of MCM, the primers oligo5 and oligo6 were used to amplify the fragment Ser499-Leu525 and the primers oligo3 and oligo2 were used to amplify the C-terminal fragment of MCM (Supplementary Table 1). The full-length *Pab*MCM was reconstructed in three PCR reactions (Fig. 1). In each reaction, the PCR to join the fragments was performed in two steps. In the first step, two fragments were incubated in the reaction mixture for 5 cycles, in absence of primers, in order to fill the 5′ ends. In the second step, primers were added and PCR was performed for 35 cycles. In the first PCR, the N-terminal fragment was fused with the fragment Ser499-Leu525 and then amplified by the Oligo1 and Oligo6 pair. In the second PCR, the fragment Ser499-Leu525 was fused with the C-terminal fragment of MCM and then amplified by the pair of oligos 5 and 2. The reconstructed full-length MCM fragment was cloned into a modified pET vector and the sequence was confirmed by sequencing.

Protein overexpression and purification. *Pab*MCM was overexpressed in Rosetta (DE3) pLysS grown in terrific broth (TB). Cells were grown to $Abs_{600} = 0.4–0.6$ before induction with 0.2 mM IPTG for 16 hours at 30 °C. *Pab*MCM was purified by heat–denaturation (20′ at 70 °C), followed by nickel affinity chromatography and size-exclusion chromatography. Fractions eluted from the nickel affinity purification step were pooled and tested for DNA contaminations by measuring the $Abs_{260}/_{280}$ ratio. Only fractions with $Abs_{260}/_{280}$ ratio between 0.51 and 0.68 (99% protein) were concentrated by ultra-filtration using a 15 R Vivaspin (30,000 MWCO), to a final concentration of 15 mg/ml. Finally, the sample was loaded onto a Superose6^{TM}10/300 GL size-exclusion chromatography column. The apparent molecular weight of PabMCM, estimated as in Andrews[24], was ~575 kDa.

DNA binding assay. DNA binding assays were carried out as described in ref. 25. Oligos 3 and 4 were annealed in annealing buffer (50 mM Tris, 50 mM NaCl, pH 8.8) to a final concentration of 10 μM. Annealing was achieved by incubating the tube containing the oligo mixture in a beaker with boiling water and letting it cool overnight. Reactions were carried out at 50 °C for 30 minutes. Annealing was checked by electrophoresis onto 1.5% agarose gels.

Helicase assay. The fluorescent-labeled Y-shaped substrate is obtained by mixing oligos A and B (Supplementary Table 2) as described in DNA binding assay and followed by PAGE purification.

The helicase assays (20 μl) were performed by mixing 5 nM fluorescent-labeled Y-shaped substrate, 10 mM ATP, and increasing amounts of enzyme in reaction buffer containing 20 mM Tris-acetate pH 7.9, 50 mM potassium-acetate, 10 mM magnesium-acetate, and 1 mM DTT. Reactions were initiated by the addition of ATP, and the mixtures were incubated at 65 °C for 45 minutes. Reactions were terminated by the addition of 11 μl of chilled stop solution (0.5 mg/ml Protéinase K, 0.5% [wt/vol] SDS, 5 mM EDTA, and 250-fold molar excess of unlabeled DNA trap oligonucleotide C to minimize reannealing of the unwound oligonucleotides. Followed by incubation at room temperature for 15 minutes, 4 μl loading buffer are added to the samples. As a positive control, DNA substrates were heat denatured at 95 °C for 15 min in the absence of helicase. As a negative control, the substrates were incubated in the reaction mix in the absence of helicase. Samples were resolved in a 8% (vol/vol) polyacrylamide gel in 1 × Tris-borate-EDTA. The fluorescent-labeled bands were visualized using a Typhoon 9400 (GE healthcare) and quantified with ImageQuant 5.2 software (Molecular Dynamics).

Electron Microscopy. *Pab*MCM was studied both by negative staining and cryo electron microscopy (EM) and single particle analysis (SPA). The protein sample was added to glow-discharged grids, blotted and washed twice before staining using 1% uranyl acetate. Data were collected on a FEI F20 FEG microscope, equipped with a TemCam-F816 (8kx8k) CCD camera. Images were collected under low dose mode at nominal magnification of 50,000x, at a final sampling of 1.51 Å/pixel at the specimen level. Negatively stained single particle images were selected interactively using the e2boxer.py program from the EMAN2 single particle analysis package and extracted into boxes. Image processing was performed using the IMAGIC-5[45] and Eman[35,46] packages. The dataset was re-sampled at 6.04 Å/pixel. ~45,000 images with homogeneous staining were band-pass filtered with a high pass cutoff of 110 Å and a low pass cutoff of 18 Å. The single particle images were analysed by Multivariate Statistical Analysis with IMAGIC-5. The dataset was subjected to successive rounds of alignment and classification in order to improve the resulting image class-averages. Three-dimensional reconstruction was performed using Imagic and Eman2 protocols. The map was deposited in the PDB with accession code EMD-3487. The cryoEM data collection was performed on Quantifoil grids covered with a thin layer of carbon, plunge-frozen into liquid ethane using a Vitrobot instrument. Data collection was performed at the same magnification of the negative staining study (50,000x), in low-dose mode. The single particle analysis was performed using the Relion software[36] according to instructions. Figures were prepared with UCSF Chimera[38] and Pymol[47].

References

1. Jenkinson, E. R. & Chong, J. P. Initiation of archaeal DNA replication. *Biochem Soc Trans* **31,** 669–673, doi: 10.1042/ (2003).
2. Pan, M., Santangelo, T. J., Li, Z., Reeve, J. N. & Kelman, Z. Thermococcus kodakarensis encodes three MCM homologs but only one is essential. *Nucleic Acids Res* **39,** 9671–9680, doi: 10.1093/nar/gkr624 (2011).
3. Walters, A. D. & Chong, J. P. An archaeal order with multiple minichromosome maintenance genes. *Microbiology* **156,** 1405–1414, doi: 10.1099/mic.0.036707-0 (2010).
4. Sakakibara, N., Kelman, L. M. & Kelman, Z. Unwinding the structure and function of the archaeal MCM helicase. *Mol Microbiol* **72,** 286–296 (2009).
5. Bell, S. D. & Botchan, M. R. The minichromosome maintenance replicative helicase. *Cold Spring Harb Perspect Biol* **5,** a012807, doi: 10.1101/cshperspect.a012807 (2013).
6. Hesketh, E. L. *et al.* DNA induces conformational changes in a recombinant human minichromosome maintenance complex. *The Journal of biological chemistry* **290,** 7973–7979, doi: 10.1074/jbc.M114.622738 (2015).
7. Li, N. *et al.* Structure of the eukaryotic MCM complex at 3.8 A. *Nature* **524,** 186–191, doi: 10.1038/nature14685 (2015).
8. Barry, E. R. & Bell, S. D. DNA replication in the archaea. *Microbiol Mol Biol Rev* **70,** 876–887, doi: 10.1128/MMBR.00029-06 (2006).
9. Chong, J. P., Hayashi, M. K., Simon, M. N., Xu, R. M. & Stillman, B. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 1530–1535, doi: 10.1073/pnas.030539597 (2000).
10. Remus, D. *et al.* Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139,** 719–730, doi: 10.1016/j.cell.2009.10.015 (2009).
11. Yu, X. *et al.* The Methanobacterium thermoautotrophicum MCM protein can form heptameric rings. *EMBO Rep* **3,** 792–797, doi: 10.1093/embo-reports/kvf160 (2002).
12. Pape, T. *et al.* Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO Rep* **4,** 1079–1083, doi: 10.1038/sj.embor.embor7400010 (2003).
13. Costa, A. *et al.* Structural basis of the Methanothermobacter thermautotrophicus MCM helicase activity. *Nucleic Acids Res* **34,** 5829–5838, doi: 10.1093/nar/gkl708 (2006).

14. Chen, Y. J. *et al.* Structural polymorphism of Methanothermobacter thermautotrophicus MCM. *J Mol Biol* **346,** 389–394, doi: 10.1016/j.jmb.2004.11.076 (2005).
15. Gomez-Llorente, Y., Fletcher, R. J., Chen, X. S., Carazo, J. M. & San Martin, C. Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from Methanobacterium thermoautotrophicum. *The Journal of biological chemistry* **280,** 40909–40915, doi: 10.1074/jbc.M509760200 (2005).
16. Bae, B. *et al.* Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure* **17,** 211–222, doi: 10.1016/j.str.2008.11.010 (2009).
17. Slaymaker, I. M. *et al.* Mini-chromosome maintenance complexes form a filament to remodel DNA structure and topology. *Nucleic Acids Res* **41,** 3446–3456, doi: 10.1093/nar/gkt022 (2013).
18. Shin, J. H., Heo, G. Y. & Kelman, Z. The Methanothermobacter thermautotrophicus MCM helicase is active as a hexameric ring. *The Journal of biological chemistry* **284,** 540–546, doi: 10.1074/jbc.M806803200 (2009).
19. Miller, J. M., Arachea, B. T., Epling, L. B. & Enemark, E. J. Analysis of the crystal structure of an active MCM hexamer. *Elife* **3,** e03433, doi: 10.7554/eLife.03433 (2014).
20. Barry, E. R., Lovett, J. E., Costa, A., Lea, S. M. & Bell, S. D. Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 1051–1056, doi: 10.1073/pnas.0809192106 (2009).
21. Sakakibara, N., Kelman, L. M. & Kelman, Z. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochem Soc Trans* **37,** 7–11, doi: 10.1042/BST0370007 (2009).
22. Kasiviswanathan, R., Shin, J. H., Melamud, E. & Kelman, Z. Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains. *The Journal of biological chemistry* **279,** 28358–28366, doi: 10.1074/jbc.M403202200 (2004).
23. Fletcher, R. J., Shen, J., Holden, L. G. & Chen, X. S. Identification of amino acids important for the biochemical activity of Methanothermobacter thermautotrophicus MCM. *Biochemistry* **47,** 9981–9986, doi: 10.1021/bi800032t (2008).
24. Andrews, P. Estimation of molecular size and molecular weights of biological compounds by gel filtration. *Methods Biochem Anal* **18,** 1–53 (1970).
25. Brewster, A. S. *et al.* Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA + hexameric helicase. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 20191–20196, doi: 10.1073/pnas.0808037105 (2008).
26. Fletcher, R. J. *et al.* The structure and function of MCM from archaeal M. Thermoautotrophicum. *Nature structural biology* **10,** 160–167, doi: 10.1038/nsb893 (2003).
27. Grainge, I. *et al.* Biochemical analysis of a DNA replication origin in the archaeon Aeropyrum pernix. *J Mol Biol* **363,** 355–369, doi: 10.1016/j.jmb.2006.07.076 (2006).
28. McGeoch, A. T., Trakselis, M. A., Laskey, R. A. & Bell, S. D. Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nat Struct Mol Biol* **12,** 756–762, doi: 10.1038/nsmb974 (2005).
29. Liew, L. P. & Bell, S. D. The interplay of DNA binding, ATP hydrolysis and helicase activities of the archaeal MCM helicase. *Biochem J* **436,** 409–414, doi: 10.1042/BJ20110084 (2011).
30. Shin, J. H., Santangelo, T. J., Xie, Y., Reeve, J. N. & Kelman, Z. Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors. *The Journal of biological chemistry* **282,** 4908–4915, doi: 10.1074/jbc.M606847200 (2007).
31. Haugland, G. T., Rollor, C. R., Birkeland, N. K. & Kelman, Z. Biochemical characterization of the minichromosome maintenance protein from the archaeon Thermoplasma acidophilum. *Extremophiles* **13,** 81–88, doi: 10.1007/s00792-008-0198-y (2009).
32. White, H. E., Saibil, H. R., Ignatiou, A. & Orlova, E. V. Recognition and separation of single particles with size variation by statistical analysis of their images. *J Mol Biol* **336,** 453–460 (2004).
33. Morris, E. P. *et al.* Evidence for a remodelling of DNA-PK upon autophosphorylation from electron microscopy studies. *Nucleic Acids Res* **39,** 5757–5767, doi: 10.1093/nar/gkr146 (2011).
34. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157,** 38–46, doi: 10.1016/j.jsb.2006.05.009 (2007).
35. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157,** 38–46 (2007).
36. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180,** 519–530, doi: 10.1016/j.jsb.2012.09.006 (2012).
37. Wiedemann, C., Ohlenschlager, O., Medagli, B., Onesti, S. & Gorlach, M. (2014).
38. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J Struct Biol* **157,** 281–287, doi: 10.1016/j.jsb.2006.06.010 (2007).
39. Wriggers, W., Milligan, R. A. & McCammon, J. A. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *Journal of structural biology* **125,** 185–195 (1999).
40. Schrodinger, L. L. C. *The PyMOL Molecular Graphics System, Version* 1.8 (2015).
41. Kinebuchi, T. *et al.* Structural basis for octameric ring formation and DNA interaction of the human homologous-pairing protein Dmc1. *Mol Cell* **14,** 363–374, doi: Doi 10.1016/S1097-2765(04)00218-7 (2004).
42. Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F. & Ishino, Y. The GINS complex from Pyrococcus furiosus stimulates the MCM helicase activity. *The Journal of biological chemistry* **283,** 1601–1609, doi: 10.1074/jbc.M707654200 (2008).
43. Ogino, H. *et al.* Activation of the MCM helicase from the thermophilic archaeon, Thermoplasma acidophilum by interactions with GINS and Cdc6-2. *Extremophiles* **18,** 915–924, doi: 10.1007/s00792-014-0673-6 (2014).
44. Xia, Y. S. *et al.* The Helicase Activity of Hyperthermophilic Archaeal MCM is Enhanced at High Temperatures by Lysine Methylation. *Front Microbiol* **6,** doi: Artn 1247 10.3389/Fmicb.2015.01247 (2015).
45. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J Struct Biol* **116,** 17–24, doi: 10.1006/jsbi.1996.0004 (1996).
46. Ludtke, S. J. 3-D structures of macromolecules using single-particle analysis in EMAN. *Methods in molecular biology* **673,** 157–173, doi: 10.1007/978-1-60761-842-3_9 (2010).
47. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
48. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25,** 1605–1612 (2004).

Acknowledgements

This work was funded by a grant from the Biotechnology and Biological Sciences Research Council (BB/J005673/1 to LS). French National Research Agency [ANR-10-JCJC-1501-01] financially supported G.H. and A.P. The Electron Microscopy Facility at Edinburgh is supported by the Scottish Alliance for Life Sciences and the Wellcome Trust (WT087658MA). GC acknowledges support from the Darwin Trust of Edinburgh. We thank Patrick Forterre for the kind gift of *Pyrococcus abyssi* genomic DNA and Stephen Bell for the gift of *Sso*MCM purified protein.

Author Contributions

L.S., G.H., A.P. and G.C. designed the experiments. G.C. performed biochemical experiments and collected the EM data. G.C., S.V. and L.S. performed image processing experiments. A.P. performed the helicase assays. L.S. and G.C. wrote the main manuscript text. L.S., G.H., A.P and G.C. prepared figures. L.S., G.H., S.V. and G.C. contributed to writing and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Cannone, G. *et al.* Structure of an octameric form of the minichromosome maintenance protein from the archaeon *Pyrococcus abyssi*. *Sci. Rep.* **7**, 42019; doi: 10.1038/srep42019 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.